Journal of Computational Physics **181,** 186–221 (2002)
doi:10.1006/jcph.2002.7118

Nodal High-Order Methods on Unstructured Grids

I. Time-Domain Solution of Maxwell's Equations

J. S. Hesthaven and T. Warburton

Division of Applied Mathematics, Brown University, Box F, Providence, Rhode Island 02912
E-mail: Jan.Hesthaven@brown.edu; timw@cfm.brown.edu

Received February 2, 2001; revised May 23, 2002

We present a convergent high-order accurate scheme for the solution of linear conservation laws in geometrically complex domains. As our main example we include a detailed development and analysis of a scheme for the time-domain solution of Maxwell's equations in a three-dimensional domain. The fully unstructured spatial discretization is made possible by the use of a high-order nodal basis, employing multivariate Lagrange polynomials defined on the triangles and tetrahedra, while the equations themselves are satisfied in a discontinuous Galerkin form with the boundary conditions being enforced weakly through a penalty term. Accuracy, stability, and convergence of the semidiscrete approximation to Maxwell's equations is established rigorously and bounds on the growth of the global divergence error are provided. Concerns related to efficient implementations are discussed in detail. This sets the stage for the presentation of examples, verifying the theoretical results, and illustrating the versatility, flexibility, and robustness when solving two- and three-dimensional benchmark problems in computational electromagnetics. Pure scattering as well as penetration is discussed and high parallel performance of the scheme is demonstrated.

Key Words: high-order accuracy; spectral methods; stability; convergence; unstructured grids; Maxwell's equations; linear conservation laws.

1. INTRODUCTION

The ability to accurately and reliably model wave-dominated problems continues to be an essential technology for the development and analysis of emerging technologies such as stealth technology, noise reduction, subsurface exploration, and optical communication to name a few. These are all problems characterized by large size in terms of a characteristic

0021-9991/02 $35.00

wavelength, geometrical complexity, often being composed of a heterogeneous collection of different materials, and all requiring a high-fidelity solution with a rigorous control of the numerical errors. Even for linear problems, such requirements force one to look beyond standard computational techniques and seek new computational frameworks to enable the accurate, efficient, and robust modeling of wave phenomena over long times in geometrically complex domains.

The requirement that one can accurately propagate waves over many periods of time suggests that high-order methods should be considered [1]. On the other hand, the use of such methods is traditionally in conflict with the need for significant geometric flexibility by being restricted to fairly simple geometries. The standard approach to overcome this restriction is to introduce a multielement formulation in which the basic building block is parametrically mapped cubes in the spirit of finite element methods. This approach has been very successfully applied to the solution of problems in fluid mechanics [2–4], gasdynamics [5–10], and electromagnetics [11–15].

When applicable, such techniques are powerful, although they do suffer from the need to tile the computational domain using only hexahedral elements. Unfortunately, automated grid generation using only such elements for general three-dimensional computational problems of a realistic complexity remains a nontrivial task and is typically rather time-consuming. Furthermore, spatial adaptation, while certainly possible, is quite a challenge with a method based solely on hexahedral elements. In contrast to this, automated grid generation employing a fully unstructured grid based on tetrahedral elements is significantly more mature, due mainly to extensive developments within the finite-element community. Spatial grid adaptation is also considerably easier in such a formulation.

It is with these issues in mind that we present a computational framework that combines the strengths of a high-order formulation with the flexibility of a fully unstructured grid based on tetrahedral elements. The formulation relies on the resolution of two central issues. On one hand we shall discuss how to represent functions defined on triangles and tetrahedra to high accuracy and how this translates into the basic operations needed to solve partial differential equations. On the other hand we need to address the question of how to use such a high-order representation to formulate a convergent scheme suitable for solving systems of linear hyperbolic problems.

Much in the spirit of the original work on spectral element methods [2, 3] we focus on the formulation of efficient and flexible unstructured grid methods using nodal elements. This is in contrast to past attempts to develop high-order unstructured grid methods, which mainly have focused on the use of high-order modal expansions, e.g., [16–21]. In these works, modal expansions of almost orthogonal polynomials defined on the simplex are utilized while a straightforward monomial basis is used in [22] (see also [23] and references therein) much in the tradition of classical high-order finite element methods for elliptic problems [24, 25].

In contrast to the classical spectral element approach, however, we do not seek a globally continuous solution but rather require that the equations be satisfied in a discontinuous Galerkin/penalty fashion. This is related to the classic discontinuous Galerkin finite element method [23], although the present approach represents a more general formulation, containing the classic discontinuous Galerkin formulation as a special case. Such more general techniques are known in the context of spectral methods as penalty methods [26], and recently stable formulations on general one-dimensional [27], triangular [28], and tetrahedral domains [29] have been discussed. These methods all share the advantage of decoupling

all the elements, hence enabling high parallel efficiency, by allowing for discontinuous solutions between elements in a natural way.

While the majority of what we shall discuss is of a general nature we discuss in detail the development and analysis of a high-order accurate unstructured grid method for the solution of Maxwell's equations in the time domain. This is not only a challenging problem but also a problem of significant contemporary interest due to emerging technologies such as broad-band target illumination and penetration, advanced materials, and diffraction-based modern optics. Furthermore, Maxwell's equations serve as an excellent example of numerous other linear hyperbolic systems of equations, e.g., elasticity, acoustics, and solid mechanics, for which the proposed framework can be adapted with little effort. In part II of this work [30] we shall discuss generalizations of the proposed computational framework to allow the accurate and robust solution of systems of nonlinear conservation laws.

What remains of the paper is organized as follows. In Section 2 we set the stage by briefly describing Maxwell's equations and the scattered field formulations, as well as boundary conditions at material interfaces and metallic boundaries. The first step in the construction of the high-order unstructured grid scheme is taken in Section 3 where we introduce a Lagrangian high-order basis on a general curvilinear simplex. In the Appendix we include a discussion of techniques that allow an efficient and accurate implementation of the basic operations. By providing the basic building block for the spatial approximation, this sets the stage for the formulation of a high-order convergent scheme for solving Maxwell's equations as discussed in Section 4. The convergence of the scheme is established in the classic way through consistency and stability. A stronger and optimal result is furthermore established by showing the scheme to be error-bounded, guaranteeing at most linear growth in time and control over the growth rate. This result also enables bounds on the growth of the divergence error. Verification and performance of the complete scheme is the topic of Section 5 where we present a number of simple tests to verify the theoretical results and illustrate the efficiency, versatility, and robustness of the computational framework for the solution of two- and three-dimensional scattering and penetration problems. We shall also briefly discuss measures taken in the implementation of the scheme to ensure efficient execution on large-scale contemporary parallel computing platforms. In Section 6 we conclude by offering a few remarks and guidelines for future work.

2. THE PHYSICAL SETTING AND MAXWELL'S EQUATIONS

We shall concern ourselves with the time-domain solution of Maxwell's equations in normalized differential form

$$\frac{\partial \boldsymbol{D}}{\partial t} = \nabla \times \boldsymbol{H} + \boldsymbol{J}, \quad \frac{\partial \boldsymbol{B}}{\partial t} = -\nabla \times \boldsymbol{E}, \tag{1}$$

$$\nabla \cdot \boldsymbol{D} = \rho, \quad \nabla \cdot \boldsymbol{B} = 0, \tag{2}$$

within the general three-dimensional domain, Ω, with the charge distribution, $\rho(\boldsymbol{x}, t)$. The electric field, $\boldsymbol{E}(\boldsymbol{x}, t)$, and the electric flux density, $\boldsymbol{D}(\boldsymbol{x}, t)$, as well as the magnetic field, $\boldsymbol{H}(\boldsymbol{x}, t)$, and the magnetic flux density, $\boldsymbol{B}(\boldsymbol{x}, t)$, are related through the constitutive

relations

$$\boldsymbol{D} = \underline{\underline{\varepsilon}}_{\mathrm{r}} \boldsymbol{E}, \quad \boldsymbol{B} = \underline{\underline{\mu}}_{\mathrm{r}} \boldsymbol{H}.$$

The normalized relative permittivity tensor, $\underline{\underline{\varepsilon}}_{\mathrm{r}}$, as well as the normalized relative permeability tensor, $\underline{\underline{\mu}}_{\mathrm{r}}$, are in general anisotropic and may depend on space and time as well as the strength of the fields themselves. The current, $\boldsymbol{J}$, is typically assumed to be related to the electric field, $\boldsymbol{E}$, through Ohm's law, $\boldsymbol{J} = \sigma \boldsymbol{E}$, where σ measures the finite conductivity.

This normalized version of Maxwell's equations is recovered by introducing the normalized quantities

$$\boldsymbol{x} = \frac{\tilde{\boldsymbol{x}}}{\tilde{L}}, \quad t = \frac{\tilde{t}}{\tilde{L}/\tilde{c}_0},$$

where $\tilde{L}$ is a reference length, and $\tilde{c}_0 = (\tilde{\varepsilon}_0 \tilde{\mu}_0)^{-1/2}$ represents the dimensional vacuum speed of light. The fields themselves are normalized as

$$\boldsymbol{E} = \frac{\tilde{Z}_0^{-1} \tilde{\boldsymbol{E}}}{\tilde{H}_0}, \quad \boldsymbol{H} = \frac{\tilde{\boldsymbol{H}}}{\tilde{H}_0}, \quad \boldsymbol{J} = \frac{\tilde{\boldsymbol{J}} \tilde{L}}{\tilde{H}_0},$$

where $\tilde{Z}_0 = \sqrt{\tilde{\mu}_0/\tilde{\varepsilon}_0}$ refers to the dimensional free space intrinsic impedance, and $\tilde{H}_0$ is a dimensional reference magnetic field strength.

For simplicity, we restrict the attention to materials that can be assumed isotropic, linear, and time-invariant, with constitutive relations of the form

$$\boldsymbol{D} = \varepsilon_{\mathrm{r}} \boldsymbol{E}, \quad \boldsymbol{B} = \mu_{\mathrm{r}} \boldsymbol{H}.$$

Here $\varepsilon_{\mathrm{r}}(\boldsymbol{x})$ and $\mu_{\mathrm{r}}(\boldsymbol{x})$ refer to the relative permittivity and permeability, respectively, of the materials.

Taking the divergence of Eq. (1) and applying Eq. (2) in combination with Gauss's law for charge conservation immediately confirms that if the initial conditions satisfy Eq. (2), and the fields are evolved according to Maxwell's equations, Eq. (1), the solution will satisfy Eq. (2) at all times. Hence, one can view Eq. (2) as a consistency condition on the initial conditions and limit the solution to the time-dependent part of Maxwell's equations, Eq. (1).

With this normalization Eq. (1) takes the nondimensional form

$$\varepsilon_{\mathrm{r}} \frac{\partial \boldsymbol{E}}{\partial t} = \nabla \times \boldsymbol{H} + \boldsymbol{J}, \quad \mu_{\mathrm{r}} \frac{\partial \boldsymbol{H}}{\partial t} = -\nabla \times \boldsymbol{E}, \tag{3}$$

which is the general form of the equations we consider.

To solve Maxwell's equations in the vicinity of boundaries, penetrable or not, we shall need boundary conditions relating the field components on either side of the boundary. Assuming that a normal unit vector, $\hat{\boldsymbol{n}}$, to the boundary is given, the boundary conditions take the form

$$\hat{\boldsymbol{n}} \times (\boldsymbol{E}_1 - \boldsymbol{E}_2) = 0, \quad \hat{\boldsymbol{n}} \times (\boldsymbol{H}_1 - \boldsymbol{H}_2) = 0, \tag{4}$$

assuming that the materials have finite conductivity; i.e., the tangential components are continuous regardless of the materials.

For the important special case of a perfect conductor, the conditions become

$$\hat{\boldsymbol{n}} \times \boldsymbol{E} = 0, \quad \hat{\boldsymbol{n}} \cdot \boldsymbol{B} = 0, \tag{5}$$

as the perfect conductor supports surface charges and currents while the fields are unable to penetrate into the body.

2.1. *The Scattered Field Formulation*

For problems involving linear materials it is often advantageous to exploit the linearity of Maxwell's equations and solve for the scattered field, $(\boldsymbol{E}^{\mathrm{s}}, \boldsymbol{H}^{\mathrm{s}})$, rather than for the total field, $(\boldsymbol{E}, \boldsymbol{H})$. These are trivially related as

$$\boldsymbol{E} = \boldsymbol{E}^{\mathrm{i}} + \boldsymbol{E}^{\mathrm{s}}, \quad \boldsymbol{H} = \boldsymbol{H}^{\mathrm{i}} + \boldsymbol{H}^{\mathrm{s}},$$

where $(\boldsymbol{E}^{\mathrm{i}}, \boldsymbol{H}^{\mathrm{i}})$ represents the incident field, illuminating the scattering object. Assuming that $(\boldsymbol{E}^{\mathrm{i}}, \boldsymbol{H}^{\mathrm{i}})$ represents a particular solution to Maxwell's equations, one recovers the scattered field formulation

$$\varepsilon_{\mathrm{r}} \frac{\partial \boldsymbol{E}^{\mathrm{s}}}{\partial t} = \nabla \times \boldsymbol{H}^{\mathrm{s}} + \sigma \boldsymbol{E}^{\mathrm{s}} - \left(\varepsilon_{\mathrm{r}} - \varepsilon_{\mathrm{r}}^{\mathrm{i}}\right) \frac{\partial \boldsymbol{E}^{\mathrm{i}}}{\partial t} + (\sigma - \sigma^{i}) \boldsymbol{E}^{\mathrm{i}}, \tag{6}$$

$$\mu_{\mathrm{r}} \frac{\partial \boldsymbol{H}^{\mathrm{s}}}{\partial t} = -\nabla \times \boldsymbol{E}^{\mathrm{s}} - \left(\mu_{\mathrm{r}} - \mu_{\mathrm{r}}^{\mathrm{i}}\right) \frac{\partial \boldsymbol{H}^{\mathrm{i}}}{\partial t}, \tag{7}$$

where $\varepsilon_{\mathrm{r}}^{\mathrm{i}}(\boldsymbol{x})$, $\mu_{\mathrm{r}}^{\mathrm{i}}(\boldsymbol{x})$, and $\sigma^{\mathrm{i}}(\boldsymbol{x})$ refer to the relative permittivity, permeability, and conductivity of the media in which the incident field represents a solution to Maxwell's equations. To simplify matters we have assumed Ohm's law, $\boldsymbol{J} = \sigma \boldsymbol{E}$. We note that the important special case of a vacuum field illuminating the scattering object is obtained by using $\varepsilon_{\mathrm{r}}^{\mathrm{i}} = \mu_{\mathrm{r}}^{\mathrm{i}} = 1$, $\sigma^{\mathrm{i}} = 0$, and using a free space solution in the forcing function.

In this formulation, the boundary conditions along a dielectric interface are

$$\hat{\boldsymbol{n}} \times \left(\boldsymbol{E}_1^{\mathrm{s}} - \boldsymbol{E}_2^{\mathrm{s}}\right) = 0, \quad \hat{\boldsymbol{n}} \times \left(\boldsymbol{H}_1^{\mathrm{s}} - \boldsymbol{H}_2^{\mathrm{s}}\right) = 0, \tag{8}$$

for the tangential components, while the conditions on the scattered field components become

$$\hat{\boldsymbol{n}} \times \boldsymbol{E}^{\mathrm{s}} = -\hat{\boldsymbol{n}} \times \boldsymbol{E}^{\mathrm{i}}, \quad \hat{\boldsymbol{n}} \cdot \boldsymbol{B}^{\mathrm{s}} = -\mu_{\mathrm{r}}^{\mathrm{i}} \hat{\boldsymbol{n}} \cdot \boldsymbol{H}^{\mathrm{i}}, \tag{9}$$

in the case of a perfectly conducting boundary.

3. THE NODAL ELEMENT

We shall seek approximate solutions to Maxwell's equations in a general domain, Ω, possibly containing a heterogeneous collection of scattering and penetrable bodies. To facilitate the required geometric flexibility, we represent the computational domain as the union of K nonoverlapping body-conforming d-simplices, D.

While this multielement formulation is essential to ensure geometric flexibility, it also introduces new complications, the resolution of which are at the heart of the construction

of the scheme. In particular, the use of simplices requires an understanding of how to construct high-order polynomial representations of solutions and basic operations on such elements. These are issues we shall deal with in the following. For continuity we postpone the important discussion of practical techniques for the efficient and accurate implementation of the basic operations to the Appendix.

3.1. *The Curvilinear* d-*Simplex*

We start by assuming that the computational domain, Ω, is decomposed into curvilinear d-simplices, $\mathsf{D} \subset \mathsf{R}^d$, as illustrated in Fig. 1 by a 3-simplex, a tetrahedron. For generality we shall limit much of the discussions to the three-dimensional case and regard the two-dimensional problem as a special case.

While we shall not eventually require that the faces of the tetrahedron are planar, such an assumption will, as we shall see shortly, significantly simplify matters in terms of analysis as well as implementation. It should also be noted that for most computational problems, the vast majority of the elements will have planar faces; i.e., it remains the single most important special case.

Let us introduce the standard tetrahedron, $\mathsf{I} \subset \mathsf{R}^3$, given by the vertices

$$v_{\mathrm{I}} = \begin{bmatrix} -1 \\ -1 \\ -1 \end{bmatrix}, \quad v_{\mathrm{II}} = \begin{bmatrix} 1 \\ -1 \\ -1 \end{bmatrix}, \quad v_{\mathrm{III}} = \begin{bmatrix} -1 \\ 1 \\ -1 \end{bmatrix}, \quad v_{\mathrm{IV}} = \begin{bmatrix} -1 \\ -1 \\ 1 \end{bmatrix},$$

as illustrated in Fig. 1 with the corresponding vertices in D termed v_1–v_4. In general we shall name the coordinates in the physical simplex, D, as $\boldsymbol{x} = (x, y, z)$ while the coordinates, $\boldsymbol{\xi} \in \mathsf{I}$, are referred to as $\boldsymbol{\xi} = (\xi, \eta, \zeta)$.

To relate operations on D to those on I we construct a smooth and invertible mapping, $\Psi: \mathsf{I} \to \mathsf{D}$, that uniquely relates the two simplices. In the case of a general curvilinear mapping, this can be constructed directly by the use of linear transfinite blending functions. Although lengthy, expressions of these mappings are straightforwardly arrived at by

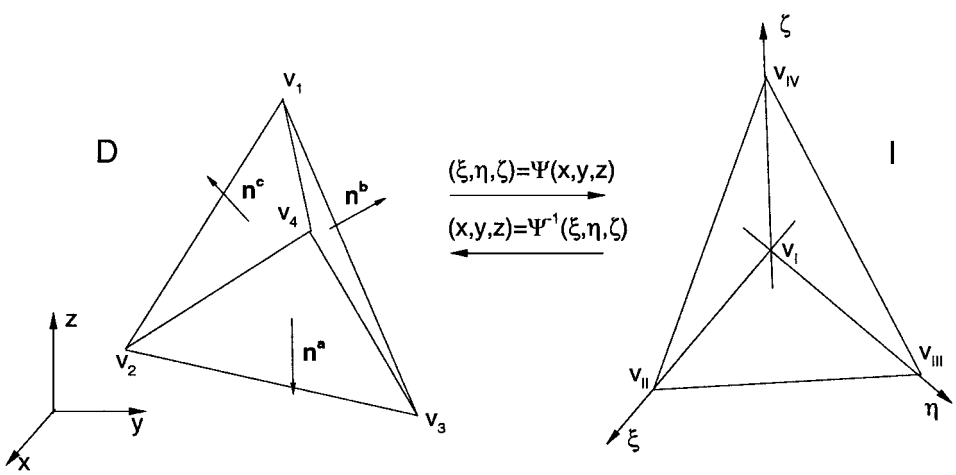

FIG. 1. Mapping between the curvilinear tetrahedral, D, and the standard tetrahedral, I, including the numbering and notation employed in the text.

blending parameterized versions of faces, edges, and the vertex coordinates. For a detailed account of this we refer to [21].

Once the mapping, $\Psi(\boldsymbol{\xi})$, has been computed we utilize it to compute the curvilinear metric of the transformation by

$$\frac{\partial \boldsymbol{x}}{\partial \boldsymbol{\xi}}\frac{\partial \boldsymbol{\xi}}{\partial \boldsymbol{x}} = \frac{\partial \Psi(\boldsymbol{\xi})}{\partial \boldsymbol{\xi}}\frac{\partial \boldsymbol{\xi}}{\partial \boldsymbol{x}} = \begin{bmatrix} x_\xi & x_\eta & x_\zeta \\ y_\xi & y_\eta & y_\zeta \\ z_\xi & z_\eta & z_\zeta \end{bmatrix}\begin{bmatrix} \xi_x & \xi_y & \xi_z \\ \eta_x & \eta_y & \zeta_z \\ \zeta_x & \zeta_y & \zeta_z \end{bmatrix} = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}.$$

This enables the computation of central operations such as gradients of scalar fields and divergence of vector fields by using standard curvilinear representations, see, e.g., [21, 29].

The metric also supplies outward pointing normal vectors at the four faces of D as

$$\begin{aligned} \boldsymbol{n}^a &= \nabla\xi + \nabla\eta + \nabla\zeta, \\ \boldsymbol{n}^b &= -\nabla\xi, \quad \boldsymbol{n}^c = -\nabla\eta, \quad \boldsymbol{n}^d = -\nabla\zeta. \end{aligned}$$

It is worthwhile recognizing that for the special case of the mapping between straightsided tetrahedra, the transformation Jacobian, J, and the full metric, $\nabla\xi, \nabla\eta$, and $\nabla\zeta$, are constant; i.e., every two straightsided tetrahedra are connected through a simple linear transformation [21, 29]. As we discuss in detail in the Appendix, this observation can be exploited to simplify the implementation of the general unstructured scheme by introducing template operators.

Let us finally define a number of different inner products on the curvilinear simplex, D. Consider the two smooth functions, $f[\mathsf{D}] \in \mathsf{C}[\mathsf{D}]$ and $g[\mathsf{D}] \in \mathsf{C}[\mathsf{D}]$ for which $f(\boldsymbol{x})\colon \mathsf{D} \to \mathsf{R}$ and $g(\boldsymbol{x})\colon \mathsf{D} \to \mathsf{R}$. The inner product, the associated L^2-norm and the inner product over the surface of D are defined as

$$(f, g)_{\mathsf{D}} = \int_{\mathsf{D}} f(\boldsymbol{x})g(\boldsymbol{x})\, d\boldsymbol{x}, \quad (f, f)_{\mathsf{D}} = \|f\|^2_{\mathsf{D}}, \quad (f, g)_{\delta\mathsf{D}} = \oint_{\delta\mathsf{D}} f(\boldsymbol{x})g(\boldsymbol{x})\, d\boldsymbol{x}.$$

These local inner products and norms form the basis for the corresponding global broken measures as

$$\begin{aligned} (f, g)_\Omega &= \sum_k (f, g)_{\mathsf{D}^k}, \quad (f, f)_\Omega = \sum_k \|f\|^2_{\mathsf{D}^k} = \|f\|^2_\Omega, \\ (f, g)_{\delta\Omega} &= \sum_k \oint_{\delta\mathsf{D}^k} f(\boldsymbol{x})g(\boldsymbol{x})\, d\boldsymbol{x}. \end{aligned}$$

3.2. *A Multivariate Polynomial Basis on the* d-*Simplex*

With the curvilinear framework in place we can now focus the attention on the development of a high-order representation of a function defined on the elemental element, I, rather than a general D.

Contrary to the approach taken in [17, 21], where a purely modal approximation is utilized, we shall use a purely nodal scheme. Hence, we assume that the unknown solutions, $\boldsymbol{q}(\boldsymbol{\xi}, t)$,

can be well approximated as

$$\boldsymbol{q}_N(\boldsymbol{\xi}, t) = \sum_{j=0}^{N} \boldsymbol{q}(\boldsymbol{\xi}_j, t) L_j(\boldsymbol{\xi}),$$

where $L_j(\boldsymbol{\xi})$ is the genuine three-dimensional multivariate Lagrange interpolation polynomial, $L_j(\boldsymbol{\xi}) \in \mathsf{P}_n^3$, where

$$\mathsf{P}_n^3 = \text{span}\{\xi^i \eta^j \zeta^k;\, i, j, k \geq 0;\, i + j + k \leq n\},$$

based on the $N_n^3 = N + 1$ nodal points, $\boldsymbol{\xi}_j$, given in the interior as well as on the boundary of I. It is easy to see that the minimum number of nodal points that will allow this basis to be complete is

$$N_n^3 = \frac{1}{6}(n+1)(n+2)(n+3),$$

where n signifies the maximum order of the polynomial.

The crucial choice of a nodal set, well suited for Lagrange interpolation within the tetrahedron, is an issue that has received some attention lately with examples of suitable nodal sets given in [31] and [29]. These both present fully unstructured nodal sets with a large degree of symmetry, exactly N_n^3 nodes within the tetrahedron and a very well-behaved Lagrange polynomial as measured through the growth of the associated Lebesque constant. Furthermore, both nodal sets include the four vertices in I and have exactly $\frac{1}{2}(n+1)(n+2)$ nodes at each of the four faces. This latter property is important as it ensures that a complete two-dimensional polynomial is supported by the nodes on each face. We use the nodal set from [29] in what remains as the nodes on which the Lagrange interpolation polynomials are based. These nodal sets are given for n up to 10, corresponding to $N_{10}^3 = 286$ nodal points within each tetrahedron and 66 nodal points at each face.

Once we have identified a proper nodal set, we can proceed with the formulation of the interpolation

$$\mathcal{I}_N^3 f(\boldsymbol{\xi}_j) = f(\boldsymbol{\xi}_j). \tag{10}$$

For the actual construction of the interpolation polynomials, let us introduce the complete polynomial basis, $p_i(\boldsymbol{\xi}) \in P_n^3$, and express the interpolation property, Eq. (10), as

$$\forall i : f(\boldsymbol{\xi}_i) = \sum_{j=0}^{N} \hat{f}_j p_j(\boldsymbol{\xi}_i) \Rightarrow \mathbf{V}\hat{\boldsymbol{f}} = \boldsymbol{f}, \tag{11}$$

where $\hat{\boldsymbol{f}} = [\hat{f}_0, \ldots, \hat{f}_N]^{\mathrm{T}}$ is the vector of expansion coefficients, $\boldsymbol{f} = [f(\boldsymbol{\xi}_0), \ldots, f(\boldsymbol{\xi}_N)]^{\mathrm{T}}$ is the grid vector, and $\mathbf{V}_{ij} = p_j(\boldsymbol{\xi}_i)$ is the multidimensional Vandermonde matrix. Clearly, for the interpolation to exist $\mathbf{V}$ must be nonsingular. Under the assumption of existence and uniqueness of the interpolation polynomial, we can express Eq. (11) as

$$\forall i : f(\boldsymbol{\xi}_i) = \sum_{j=0}^{N} f(\boldsymbol{\xi}_i) L_j(\boldsymbol{\xi}_i). \tag{12}$$

Combining Eqs. (11) and (12) implies that

$$\mathbf{V}^{\mathrm{T}}\boldsymbol{L} = \boldsymbol{p}, \tag{13}$$

where $\boldsymbol{L} = [L_0(\boldsymbol{\xi}), \ldots, L_N(\boldsymbol{\xi})]^{\mathrm{T}}$ and $\boldsymbol{p} = [p_0(\boldsymbol{\xi}), .., p_N(\boldsymbol{\xi})]^{\mathrm{T}}$. This enables the evaluation of $\mathcal{I}_N^3 f(\boldsymbol{\xi})$ anywhere in I by computing $L_j(\boldsymbol{\xi})$ and using Eq. (12).

Once the Lagrange polynomial has been computed, all other operations follow immediately, although care must be exercised to ensure accuracy and efficiency. The details of these practical, yet important, details are found in the Appendix.

4. A CONVERGENT SCHEME FOR MAXWELL'S EQUATIONS

Having realized high-order formulations of basic operations on the nodal tetrahedron, we are now in a position to develop a scheme suitable for solving linear systems of hyperbolic problems in complex geometries, exemplified by a scheme for solving Maxwell's equations.

Let us express Maxwell's equations, Eq. (3), in conservation form

$$\mathbf{Q}\frac{\partial \boldsymbol{q}}{\partial t} + \nabla \cdot \boldsymbol{F}(\boldsymbol{q}) = \boldsymbol{S}, \tag{14}$$

where the material matrix, $\mathbf{Q}(\boldsymbol{x})$, the state vector, $\boldsymbol{q}$, and the flux, $\boldsymbol{F}(\boldsymbol{q})$, are

$$\mathbf{Q}(\boldsymbol{x}) = \begin{bmatrix} \varepsilon_{\mathrm{r}} & 0 \\ 0 & \mu_{\mathrm{r}} \end{bmatrix}, \quad \boldsymbol{q} = \begin{bmatrix} \boldsymbol{E} \\ \boldsymbol{H} \end{bmatrix}, \quad F_i(\boldsymbol{q}) = \begin{bmatrix} -\boldsymbol{e}_i \times \boldsymbol{H} \\ \boldsymbol{e}_i \times \boldsymbol{E} \end{bmatrix},$$

with $\boldsymbol{F}(\boldsymbol{q}) = [F_1(\boldsymbol{q}), F_2(\boldsymbol{q}), F_3(\boldsymbol{q})]^{\mathrm{T}}$. Here $\boldsymbol{e}_i$ signifies the three Cartesian unit vectors and $\boldsymbol{S} = [\boldsymbol{S}^E, \boldsymbol{S}^H]^{\mathrm{T}}$ represents body forces, e.g., currents, and terms introduced by the scattered field formulation, Eqs. (6) and (7).

4.1. *Central Elements of the Scheme*

Let us begin by using the nodal basis discussed in the previous section and assume that the statevector, $\boldsymbol{q}$, can be represented as

$$\boldsymbol{q}_N(\boldsymbol{x}, t) = \sum_{j=0}^{N} \boldsymbol{q}(\boldsymbol{x}_j, t) L_j(\boldsymbol{x}) = \sum_{j=0}^{N} \boldsymbol{q}_j(t) L_j(\boldsymbol{x}),$$

within each general curvilinear element, D^k, with $\boldsymbol{x}_j = \boldsymbol{x}(\xi_j)$ through the curvilinear mapping discussed in Section 3.1.

We shall consider schemes in which we require Eq. (14) to be satisfied elementwise in the following way:

$$\int_{\mathsf{D}} \left(\mathbf{Q}\frac{\partial \boldsymbol{q}_N}{\partial t} + \nabla \cdot \boldsymbol{F}_N - \boldsymbol{S}_N \right) \phi_i(\boldsymbol{x})\, d\boldsymbol{x} = \oint_{\delta \mathsf{D}} \psi_i(\boldsymbol{x}) \hat{\boldsymbol{n}} \cdot (\boldsymbol{F}_N - \boldsymbol{F}_N^*)\, d\boldsymbol{x}. \tag{15}$$

Here ϕ_i and ψ_i signify sequences of N test functions while $\hat{\boldsymbol{n}}$ is an outward-pointing normal vector and $\boldsymbol{F}_N^*$ is a numerical flux, the specification of which we return to shortly.

It is worthwhile emphasizing a few characteristics of this formulation, Eq. (15). We see that consistency of the scheme depends solely on the consistency of the inner scheme and the

numerical flux. One should also observe that boundary/interface conditions are not imposed exactly but rather weakly through the penalizing surface integral. Finally we emphasize that in a multielement context, the formulation is inherently discontinuous, enforcing the interface conditions weakly through the penalizing term and giving rise to a highly parallel formulation.

In choosing ϕ_i, ψ_i, and the numerical flux, $\boldsymbol{F}_N^*$, one has a tremendous amount of freedom in designing schemes suitable for solving differential equations. In [10] we proposed stable spectral collocation methods with weakly imposed boundary/interface conditions for solving the advection–diffusion equation and the compressible Navier–Stokes equations by choosing $\phi_i(\boldsymbol{x}) = \psi_i(\boldsymbol{x}) = \delta(\boldsymbol{x} - \boldsymbol{x}_i)$. Alternative choices, likewise leading to stable schemes for solving linear conservation laws, were discussed in [28, 29]. There we considered mixed Galerkin–collocation formulations by choosing $\phi_i(\boldsymbol{x}) = L_i(\boldsymbol{x})$, as in a classic Galerkin formulation, but using $\psi_i(\boldsymbol{x}) = \delta(\boldsymbol{x} - \boldsymbol{x}_i)$ to impose the boundary/interface conditions. In all cases, the numerical flux, $\boldsymbol{F}_N^*$, was chosen to reflect upwinding, although any consistent numerical flux could be used, see, e.g., [23] for discussion of alternatives.

Here, we require that the equations, Eq. (3), be satisfied as

$$\int_{\mathsf{D}} \left(\mathbf{Q} \frac{\partial \boldsymbol{q}_N}{\partial t} + \nabla \cdot \boldsymbol{F}_N - \boldsymbol{S}_N \right) L_i(\boldsymbol{x})\, d\boldsymbol{x} = \oint_{\delta \mathsf{D}} \tau(\boldsymbol{x}) L_i(\boldsymbol{x}) \hat{\boldsymbol{n}} \cdot (\boldsymbol{F}_N - \boldsymbol{F}_N^*)\, d\boldsymbol{x}; \tag{16}$$

i.e., in the language of the general formulation in Eq. (15) we have $\phi_i(\boldsymbol{x}) = L_i(\boldsymbol{x})$ and $\psi_i(\boldsymbol{x}) = \tau L_i(\boldsymbol{x})$ where $\tau(\boldsymbol{x})$ is a free parameter to be specified later.

It is worth noting that the classical discontinuous Galerkin formulation [23] is recovered from Eq. (16) by a simple integration by parts; i.e., it is a special case of the scheme in Eq. (15).

While there are many possibilities for the numerical flux, $\boldsymbol{F}_N^*$, the linearity of the problem suggests that upwinding is a natural solution. To understand the form of this, it is helpful to recall that

$$\hat{\boldsymbol{n}} \cdot \boldsymbol{F}_N = \begin{bmatrix} -\hat{\boldsymbol{n}} \times \boldsymbol{H}_N \\ \hat{\boldsymbol{n}} \times \boldsymbol{E}_N \end{bmatrix},$$

i.e., the normal component of the fluxes represents nothing else than the tangential field components and the purpose of the right-hand side in Eq. (16) is to weakly enforce continuity of the tangential field components at the face of the elements. This yields the explicit form of the upwinded penalizing boundary term as [32]

$$\hat{\boldsymbol{n}} \cdot (\boldsymbol{F}_N - \boldsymbol{F}_N^*) = \begin{cases} \bar{Z}^{-1} \hat{\boldsymbol{n}} \times (Z^+ [\boldsymbol{H}_N] - \hat{\boldsymbol{n}} \times [\boldsymbol{E}_N]) \\ \bar{Y}^{-1} \hat{\boldsymbol{n}} \times (-\hat{\boldsymbol{n}} \times [\boldsymbol{H}_N] - Y^+ [\boldsymbol{E}_N]), \end{cases} \tag{17}$$

where

$$[\boldsymbol{E}_N] = \boldsymbol{E}_N^+ - \boldsymbol{E}_N^-, \quad [\boldsymbol{H}_N] = \boldsymbol{H}_N^+ - \boldsymbol{H}_N^-$$

measures the jump in the field values across an interface; i.e., superscript "+" refers to field values from the neighbor element while superscript "−" refers to field values local to the element. To account for the possible differences in material properties between the

two elements, we have the local impedance, $Z^{\pm}$, and conductance, $Y^{\pm}$, as

$$Z^{\pm} = \frac{1}{Y^{\pm}} = \sqrt{\frac{\mu_{\mathrm{r}}^{\pm}}{\varepsilon_{\mathrm{r}}^{\pm}}},$$

and the sums,

$$\bar{Z} = Z^{+} + Z^{-}, \quad \bar{Y} = Y^{+} + Y^{-},$$

of the local impedance and conductance, respectively.

The special case of a perfectly conducting wall is handled in the above formulation by defining a mirror state within the metallic scatterer as

$$\hat{\boldsymbol{n}} \times \boldsymbol{E}_N^{+} = -\hat{\boldsymbol{n}} \times \boldsymbol{E}_N^{-}, \quad \hat{\boldsymbol{n}} \times \boldsymbol{H}_N^{+} = -\hat{\boldsymbol{n}} \times \boldsymbol{H}_N^{-},$$

and define the material parameters as $Z^{+} = Z^{-}$.

Now returning to the semidiscrete scheme, Eq. (16), we have an elementwise expression for the electric field components

$$\sum_{j=0}^{N} \left(\mathbf{M}_{ij}^{\varepsilon} \frac{d\boldsymbol{E}_j}{dt} - \mathbf{S}_{ij} \times \boldsymbol{H}_j - M_{ij} \boldsymbol{S}_j^{E} \right) = \sum_{l} \mathbf{F}_{il} \left(\hat{\boldsymbol{n}}_l \times \frac{Z_l^{+}[\boldsymbol{H}_l] - \hat{\boldsymbol{n}}_l \times [\boldsymbol{E}_l]}{Z_l^{+} + Z_l^{-}} \right), \tag{18}$$

and for the magnetic field components

$$\sum_{j=0}^{N} \left(\mathbf{M}_{ij}^{\mu} \frac{d\boldsymbol{H}_j}{dt} + \mathbf{S}_{ij} \times \boldsymbol{E}_j - M_{ij} \boldsymbol{S}_j^{H} \right) = \sum_{l} \mathbf{F}_{il} \left(\hat{\boldsymbol{n}}_l \times \frac{-\hat{\boldsymbol{n}}_l \times [\boldsymbol{H}_l] - Y_l^{+}[\boldsymbol{E}_l]}{Y_l^{+} + Y_l^{-}} \right). \tag{19}$$

Here we have

$$\mathbf{M}_{ij}^{\varepsilon} = (L_i(\boldsymbol{x}), \varepsilon(\boldsymbol{x}) L_j(\boldsymbol{x}))_{\mathsf{D}}, \quad \mathbf{M}_{ij}^{\mu} = (L_i(\boldsymbol{x}), \mu(\boldsymbol{x}) L_j(\boldsymbol{x}))_{\mathsf{D}},$$

as the material-scaled mass matrices and

$$\mathbf{M}_{ij} = (L_i(\boldsymbol{x}), L_j(\boldsymbol{x}))_{\mathsf{D}}, \quad \mathbf{S}_{ij} = \left(\mathbf{S}_{ij}^{x}, \mathbf{S}_{ij}^{y}, \mathbf{S}_{ij}^{z} \right) = (L_i(\boldsymbol{x}), \nabla L_i(\boldsymbol{x}))_{\mathsf{D}} \tag{20}$$

represent the local mass and stiffness matrices. Note that in the special case where ε_{r} and μ_{r} are elementwise constant, we recover $(\mathbf{M}^{\varepsilon}, \mathbf{M}^{\mu}) = (\varepsilon_{\mathrm{r}}\mathbf{M}, \mu_{\mathrm{r}}\mathbf{M})$.

We have, furthermore, introduced the face-based mass matrices

$$\mathbf{F}_{il} = (L_i(\boldsymbol{x}), \tau(\boldsymbol{x}) L_l(\boldsymbol{x}))_{\delta\mathsf{D}},$$

where the second index is limited to the trace of the nodal set at the faces of D.

Expressing Eqs. (18) and (19) in fully explicit form yields

$$\frac{d\boldsymbol{E}_N}{dt} = (\mathbf{M}^{\varepsilon})^{-1} \mathbf{S} \times \boldsymbol{H}_N + (\mathbf{M}^{\varepsilon})^{-1} \mathbf{M} \boldsymbol{S}^{E} + (\mathbf{M}^{\varepsilon})^{-1} \mathbf{F} \left(\hat{\boldsymbol{n}} \times \frac{Z^{+}[\boldsymbol{H}_N] - \hat{\boldsymbol{n}} \times [\boldsymbol{E}_N]}{Z^{+} + Z^{-}} \right) \Bigg|_{\delta\mathsf{D}}, \tag{21}$$

and

$$\frac{d\boldsymbol{H}_N}{dt} = -(\mathbf{M}^{\mu})^{-1}\mathbf{S}\times\boldsymbol{E}_N + (\mathbf{M}^{\mu})^{-1}\mathbf{M}\boldsymbol{S}^H - (\mathbf{M}^{\mu})^{-1}\mathbf{F}\left(\hat{\boldsymbol{n}}\times\frac{\hat{\boldsymbol{n}}\times[\boldsymbol{H}_N]+Y^+[\boldsymbol{E}_N]}{Y^+ + y^-}\right)\Bigg|_{\delta\mathsf{D}}. \tag{22}$$

The mass matrices, $\mathbf{M}$ and $\mathbf{M}^{\varepsilon,\mu}$, can be computed exactly as described in the Appendix and inverted straightforwardly. We also need to initialize

$$(\mathbf{M}^{\varepsilon,\mu})^{-1}\mathbf{S} = (\mathbf{M}^{\varepsilon,\mu})^{-1}[\mathbf{S}^x, \mathbf{S}^y, \mathbf{S}^z]^{\mathrm{T}},$$

which represents the general curvilinear gradient operator matrix, as well as $(\mathbf{M}^{\varepsilon,\mu})^{-1}\mathbf{M}$ for the source terms. It is worth noting that for all straightfaced tetrahedra with constant material parameters, the entries of $\mathbf{S}$ can be formed directly from template matrices, defined for the standard element, by simple linear combinations. The same holds true for the face-based operators $\mathbf{M}^{-1}\mathbf{F}$, which likewise can be precomputed for all straightfaced elements with constant materials by linear scaling from standard template operators. An individual initialization is only required for general curved elements and elements with smoothly varying material parameters.

4.2. *Consistency*

To set the stage for the analysis of the scheme, let us introduce the exact solution, $\boldsymbol{q} = [\boldsymbol{E}, \boldsymbol{H}]$, to Maxwell's equations, Eq. (3), as well as its projection, $\mathcal{P}_N\boldsymbol{q} = [\mathcal{P}_N\boldsymbol{E}, \mathcal{P}_N\boldsymbol{H}]^{\mathrm{T}}$, on the space spanned by n-order polynomials, i.e., $\mathcal{P}_N\boldsymbol{q} \in \mathsf{P}_n^3$. Except in very special cases $\mathcal{P}_N\boldsymbol{q}$ will generally be different from the numerical solution, $\boldsymbol{q}_N = [\boldsymbol{E}_N, \boldsymbol{H}_N]^{\mathrm{T}}$, which is the exact solution to the discrete problem, Eqs. (21)–(22).

In the following we shall assume that $\mathcal{P}_N$ represents the L^2-projection, although in practice it is in fact an interpolation operator. A subtle difference between these two representations is the possibility of a discrete aliasing error through the interpolation of the initial conditions. One could avoid this by reading the nodal values of the Galerkin projection of the initial conditions, computed by using a quadrature of sufficiently high order. However, if the initial conditions are smooth and well resolved, this discrete aliasing error can be assumed to be small. We shall not explicitly distinguish between these two representations in what remains.

A central result throughout shall be:

LEMMA 4.1. *Assume that $u \in \boldsymbol{H}^p(\mathsf{D})$, $p \geq 0$. Then there exists a constant, C, dependent on p and the angle condition of D, but independent of u, $h = \operatorname{diam}(\mathsf{D})$, and n, such that*

$$\|u - \mathcal{P}_N u\|_{\boldsymbol{H}^q(\mathsf{D})} \leq C\frac{h^{\sigma-q}}{n^{p-2q}}\|u\|_{\boldsymbol{H}^p(\mathsf{D})},$$

where $\sigma = \min(p, n+1)$ and $0 \leq q \leq \sigma$.

Here we have introduced the standard Sobolev norm

$$\|u\|^2_{H^p(\mathsf{D})} = \sum_{|\alpha|\leq p}\left\|\frac{\partial^{\alpha_1}}{\partial x^{\alpha_1}}\frac{\partial^{\alpha_2}}{\partial x^{\alpha_2}}\frac{\partial^{\alpha_3}}{\partial x^{\alpha_3}}u\right\|^2_{\mathsf{D}},$$

with the multi-index, $\alpha = (\alpha_1, \alpha_2, \alpha_3)$ and the associated space of functions, $H^p(\mathsf{D})$, for which $\|u\|_{H^p(\mathsf{D})}$ is bounded. The result follows directly by combining classical results from polynomial approximation with results from finite element analysis for h-refinement [24, 25, 33, 34].

It is worth noting that Lemma 4.1 can possibly be improved at most $\sqrt{p}$ as that result is sharp for the one-dimensional case [25]. Furthermore, by using a projection different from L^2 one can improve further on the bound [24]. We shall, however, focus on estimates using L^2-projections for simplicity and not revisit this detail again.

Lemma 4.1 suffices to ensure consistency of the spatial discretization in Eqs. (18) and (19) provided only that the exact solution is sufficently smooth. Furthermore, for analytic fields one can expect exponentially vanishing truncation errors and, thus, errors, provided the scheme is stable. This, however, is an issue that requires more of a careful discussion.

4.3. *Stability*

Let us attend to the issue of semidiscrete stability and define the local energy

$$E^k = \frac{1}{2}\int_{\mathsf{D}^k} (\mu|\boldsymbol{H}_N|^2 + \varepsilon|\boldsymbol{E}_N|^2)\, d\boldsymbol{x},$$

and the associated global energy, $E = \sum_k E^k$.

LEMMA 4.2 (TWO-ELEMENT STABILITY). *Assume that a solution to Maxwell's equations exists on a domain consisting of two elements sharing one face. Stability of the semidiscrete approximation of Maxwell's equations*, *Eqs.* (21)–(22), *on this domain is guaranteed provided*

$$\tau = 1.$$

Proof. Consider Maxwell's equations on the semidiscrete form, Eqs. (18)–(19). Multiply from the left with $(\boldsymbol{E}_j, \boldsymbol{H}_j)$ and sum over all the nodes in D to obtain

$$\frac{1}{2}\frac{d}{dt}(\boldsymbol{E}_N^-, \varepsilon\boldsymbol{E}_N^-)_{\mathsf{D}} = (\boldsymbol{E}_N^-, \nabla\times\boldsymbol{H}_N^-)_{\mathsf{D}} + (\boldsymbol{E}_N^-, \boldsymbol{S}^E)_{\mathsf{D}} + \oint_{\delta\mathsf{D}} \tau\boldsymbol{E}_N^- \cdot \left(\hat{\boldsymbol{n}}^- \times \frac{Z^+[\boldsymbol{H}_N] - \hat{\boldsymbol{n}}^- \times [\boldsymbol{E}_N]}{Z^+ + Z^-}\right) d\boldsymbol{x},$$

and

$$\frac{1}{2}\frac{d}{dt}(\boldsymbol{H}_N^-, \mu\boldsymbol{H}_N^-)_{\mathsf{D}} = -(\boldsymbol{H}_N^-, \nabla\times\boldsymbol{E}_N^-)_{\mathsf{D}} + (\boldsymbol{H}_N^-, \boldsymbol{S}^H)_{\mathsf{D}} - \oint_{\delta\mathsf{D}} \tau\boldsymbol{H}_N^- \cdot \left(\hat{\boldsymbol{n}}^- \times \frac{Y^+[\boldsymbol{E}_N] + \hat{\boldsymbol{n}}^- \times [\boldsymbol{H}_N]}{Y^+ + Y^-}\right) d\boldsymbol{x}.$$

Addition of the two contributions, and application of the divergence theorem and standard

vector identities yields

$$\frac{d}{dt}E^k = \oint_{\delta \mathsf{D}} (1-\tau)\hat{\boldsymbol{n}}^- \cdot (\boldsymbol{H}_N^- \times \boldsymbol{E}_N^-) + \tau\left(\frac{Y^+}{\bar{Y}}\boldsymbol{E}_N^- \cdot (\hat{\boldsymbol{n}}^- \times \boldsymbol{H}_N^+) - \frac{Z^+}{\bar{Z}}\boldsymbol{H}_N^- \cdot (\hat{\boldsymbol{n}}^- \times \boldsymbol{E}_N^+)\right)$$
$$- \tau\left(\frac{1}{\bar{Z}}\boldsymbol{E}_N^- \cdot (\hat{\boldsymbol{n}}^- \times (\hat{\boldsymbol{n}}^- \times [\boldsymbol{E}_N])) + \frac{1}{\bar{Y}}\boldsymbol{H}_N^- \cdot (\hat{\boldsymbol{n}}^- \times (\hat{\boldsymbol{n}}^- \times [\boldsymbol{H}_N]))\right) d\boldsymbol{x}$$
$$+ (\boldsymbol{E}_N^-, \boldsymbol{S}^E)_{\mathsf{D}} + (\boldsymbol{H}_N^-, \boldsymbol{S}^H)_{\mathsf{D}}.$$

To understand the stability of a common edge, it suffices to consider the case where $\boldsymbol{S}^E = \boldsymbol{S}^H = 0$. Adding the contribution from two edges, utilizing that $\hat{\boldsymbol{n}}^- = -\hat{\boldsymbol{n}}^+$, yields

$$\frac{d}{dt}E = \oint_{\delta \mathsf{D}} (1-\tau)(\hat{\boldsymbol{n}}^- \cdot \boldsymbol{H}_N^- \times \boldsymbol{E}_N^- - \hat{\boldsymbol{n}}^- \cdot \boldsymbol{H}_N^+ \times \boldsymbol{E}_N^+)$$
$$+ \frac{\tau}{\bar{Z}}[\boldsymbol{E}_N] \cdot \hat{\boldsymbol{n}}^- \times \hat{\boldsymbol{n}}^- \times [\boldsymbol{E}_N] + \frac{\tau}{\bar{Y}}[\boldsymbol{H}_N] \cdot \hat{\boldsymbol{n}}^- \times \hat{\boldsymbol{n}}^- \times [\boldsymbol{H}_N]\, d\boldsymbol{x}$$
$$= -\oint_{\delta \mathsf{D}} (1-\tau)\hat{\boldsymbol{n}}^- \cdot (\boldsymbol{H}_N^+ \times \boldsymbol{E}_N^+ - \boldsymbol{H}_N^- \times \boldsymbol{E}_N^-)$$
$$- \frac{\tau}{\bar{Z}}|\hat{\boldsymbol{n}}^- \times [\boldsymbol{E}_N]|^2 - \frac{\tau}{\bar{Y}}|\hat{\boldsymbol{n}}^- \times [\boldsymbol{H}_N]|^2\, d\boldsymbol{x}.$$

A sufficient condition for E to remain bounded is

$$(1-\tau)((\boldsymbol{H}_N^-)^{\mathrm{T}}\mathbf{R}\boldsymbol{E}_N^- - (\boldsymbol{H}_N^+)^{\mathrm{T}}\mathbf{R}\boldsymbol{E}_N^+) + \frac{\tau}{\bar{Z}}[\boldsymbol{E}_N]^{\mathrm{T}}\mathbf{R}^{\mathrm{T}}\mathbf{R}[\boldsymbol{E}_N] + \frac{\tau}{\bar{Y}}[\boldsymbol{H}_N]^{\mathrm{T}}\mathbf{R}^{\mathrm{T}}\mathbf{R}[\boldsymbol{H}_N] \geq 0, \tag{23}$$

where we have introduced the rotation matrix

$$\mathbf{R} = \mathbf{R}(\hat{\boldsymbol{n}}) = \begin{bmatrix} 0 & -n_z & n_y \\ n_z & 0 & -n_x \\ -n_y & n_x & 0 \end{bmatrix}.$$

Inspection reveals that by defining $\boldsymbol{q} = [\boldsymbol{E}_N^+, \boldsymbol{E}_N^-, \boldsymbol{H}_N^+, \boldsymbol{H}_N^-]^{\mathrm{T}}$, Eq. (23) may be expressed as a symmetric quadratic form; i.e., it suffices to choose τ such that all eigenvalues of $\mathbf{A}$ are nonnegative. Leaving out the lengthy and purely algebraic manipulations, we consider the resulting two sets of eigenvalues of $\mathbf{A}$ given as

$$\lambda_{1,2} = 0, \quad \lambda_{3,4} = \frac{\tau}{\bar{Z}} \pm \frac{1}{2\bar{Z}}\sqrt{4\tau^2 + \bar{Z}(\tau-1)^2},$$

and

$$\lambda_{5,6} = \frac{\tau}{\bar{Y}} \pm \frac{1}{2\bar{Y}}\sqrt{4\tau^2 + \bar{Y}(\tau-1)^2}.$$

Clearly, the choice of $\tau = 1$ is the only feasible solution that ensures stability of the upwind scheme used to connect the elements.

The special case of an element being terminated by a metallic conductor is treated by using the conditions

$$\hat{\boldsymbol{n}} \times \boldsymbol{E}_N^- = -\hat{\boldsymbol{n}} \times \boldsymbol{E}_N^+, \quad \hat{\boldsymbol{n}} \times \boldsymbol{H}_N^- = \hat{\boldsymbol{n}} \times \boldsymbol{H}_N^+,$$

and $Z^+ = Z^- = Z,\ Y^+ = Y^- = Y$.

Following the same procedure as above yields the constraint

$$(1-\tau)\boldsymbol{H}_N^{\mathrm{T}}\mathbf{R}\boldsymbol{E} + \frac{\tau}{2Z}\boldsymbol{E}_N^{\mathrm{T}}\mathbf{R}^{\mathrm{T}}\mathbf{R}\boldsymbol{E}_N \geq 0.$$

Computing the eigenvalues of the corresponding quadratic form yields two pairs of the form

$$\lambda_1 = 0, \quad \lambda_{2,3} = \frac{\tau}{Z} \pm \frac{1}{Z}\sqrt{\tau^2 + Z^2(\tau-1)^2}.$$

The only way to guarantee possitivity of the eigenvalues and hence the quadratic form is to choose $\tau = 1$. ■

With these results in place, we can now state:

THEOREM 4.1 (STABILITY). *Assume that a unique solution to Maxwell's equations exists in the general domain, Ω. Assume furthermore that the boundary of Ω is either periodic or terminated with a perfectly conducting boundary.*

Then the semidiscrete approximation to Maxwell's equations, Eqs. (21)–(22), *is globally stable in the sense that*

$$\frac{d}{dt}E \leq C\big(E + \|\boldsymbol{S}^E\|_\Omega^2 + \|\boldsymbol{S}^H\|_\Omega^2\big),$$

provided only that

$$\tau = \mathbf{1}.$$

Moreover, if $\boldsymbol{S}^E = \boldsymbol{S}^H = 0$, then $C \leq 0$.

Proof. As each face is counted only once, the result follows directly by summation over all the faces and the application of Lemma 4.2

$$\begin{aligned}\frac{d}{dt}E &\leq \sum_k (\boldsymbol{E}_N, \boldsymbol{S}^E)_{\mathsf{D}^k} + (\boldsymbol{H}_N, \boldsymbol{S}^H)_{\mathsf{D}^k} \\ &\leq C\big(E + \|\boldsymbol{S}^E\|_\Omega^2 + \|\boldsymbol{S}^H\|_\Omega^2\big),\end{aligned}$$

using that $(\boldsymbol{E}_N, \boldsymbol{S}^E)_{\mathsf{D}} \leq C(\|\boldsymbol{E}_N\|_{\mathsf{D}}^2 + \|\boldsymbol{S}^E\|_{\mathsf{D}}^2)$, $\|\boldsymbol{E}_N\|_{\mathsf{D}}^2 \leq C(\boldsymbol{E}_N, \varepsilon_r\boldsymbol{E}_N)_{\mathsf{D}}$ since $\varepsilon \geq 1$. A similar line of reasoning is applicable for $(\boldsymbol{H}_N, \boldsymbol{S}^H)_{\mathsf{D}}$ and the result on global stability follows. ■

4.4. *Convergence*

With consistency and stability in equivalent norms, convergence follows directly from the equivalence theorem with a bound on the local error

$$\varepsilon_{\mathsf{D}}(t) = \|\boldsymbol{E}(t) - \boldsymbol{E}_N(t)\|_{\mathsf{D}} + \|\boldsymbol{H}(t) - \boldsymbol{H}_N(t)\|_{\mathsf{D}},$$

of the form

$$\varepsilon_{\mathsf{D}}(t) \leq Ce^{\alpha t}\left(\varepsilon_{\mathsf{D}}(0) + \int_0^t \|\boldsymbol{T}^q(s)\|_{\mathsf{D}}\, ds\right),$$

and global convergence is hence established up to exponential growth in time as is typical for Lax-type stability results.

As it turns out, however, we can do better and recover a sharp bound for the growth in time by generalizing ideas recently introduced in the context of finite difference methods [35, 36]. To realize this, let us make the natural split of the elementwise error as

$$\begin{aligned}\varepsilon_{\mathsf{D}} &\leq (\|\boldsymbol{E}-\mathcal{P}_N\boldsymbol{E}\|_{\mathsf{D}}+\|\boldsymbol{H}-\mathcal{P}_N\boldsymbol{H}\|_{\mathsf{D}})+(\|\mathcal{P}_N\boldsymbol{E}-\boldsymbol{E}_N\|_{\mathsf{D}}+\|\mathcal{P}_N\boldsymbol{H}-\boldsymbol{H}_N\|_{\mathsf{D}})\\ &= \varepsilon^a_{\mathsf{D}}+\varepsilon^b_{\mathsf{D}},\end{aligned}$$

where $\varepsilon^a_{\mathsf{D}}$ represents the error introduced by the polynomial approximation of the exact solution while $\varepsilon^b_{\mathsf{D}}$ measures the error associated with the semidiscrete approximation of Maxwell's equations.

To bound $\varepsilon^a_{\mathsf{D}}$ we need only recall Lemma 4.1 to state:

LEMMA 4.3. *Assume that* $\boldsymbol{q}=[\boldsymbol{E},\boldsymbol{H}]^{\mathrm{T}}\in\boldsymbol{H}^p(\mathsf{D})$, $p\geq 0$. *Then there exists a constant*, C, *dependent on* p *and the angle condition of* D, *but independent of* $\boldsymbol{q}$, $h=\operatorname{diam}(\mathsf{D})$, *and* n, *such that*

$$\|\boldsymbol{q}-\mathcal{P}_N\boldsymbol{q}\|_{\mathsf{D}}\leq C\frac{h^\sigma}{n^p}\|\boldsymbol{q}\|_{H^p(\mathsf{D})},$$

where $\sigma=\min(p,n+1)$.

To arrive at a bound for $\varepsilon^b_{\mathsf{D}}$, let us first consider the projection of the truncation error, $\mathcal{P}_N\boldsymbol{T}^q=[\mathcal{P}_N\boldsymbol{T}^E,\mathcal{P}_N\boldsymbol{T}^H]^{\mathrm{T}}$, on the form

$$\begin{aligned}(L_i,\mathcal{P}_N\boldsymbol{T}^E)_{\mathsf{D}} &= (L_i,\mathcal{P}_N\nabla\times\boldsymbol{H}-\mathcal{P}_N\nabla\times\mathcal{P}_N\boldsymbol{H})_{\mathsf{D}}\\ &\quad-\frac{1}{\bar{Z}}(L_i,\hat{\boldsymbol{n}}\times(Z^+[\mathcal{P}_N\boldsymbol{H}]-\hat{\boldsymbol{n}}\times[\mathcal{P}_N\boldsymbol{E}]))_{\delta\mathsf{D}},\end{aligned}\tag{24}$$

$$\begin{aligned}(L_i,\mathcal{P}_N\boldsymbol{T}^H)_{\mathsf{D}} &= -(L_i,\mathcal{P}_N\nabla\times\boldsymbol{E}-\mathcal{P}_N\nabla\times\mathcal{P}_N\boldsymbol{E})_{\mathsf{D}}\\ &\quad-\frac{1}{\bar{Y}}(L_i,\hat{\boldsymbol{n}}\times(-Y^+[\mathcal{P}_N\boldsymbol{E}]-\hat{\boldsymbol{n}}\times[\mathcal{P}_N\boldsymbol{H}]))_{\delta\mathsf{D}}.\end{aligned}\tag{25}$$

This is derived by introducing $\mathcal{P}_N\boldsymbol{q}$ into the semidiscrete scheme, Eqs. (18)–(19), and use that $\boldsymbol{q}$ satisfies Maxwell's equations, Eq. (3).

The projection of the truncation error can be bounded by the exact solution as:

LEMMA 4.4. *Assume that* $\boldsymbol{q}=[\boldsymbol{E},\boldsymbol{H}]^{\mathrm{T}}\in H^p(\mathsf{D})$, $p\geq 2$. *Then there exists a constant*, C, *dependent on* p, *the angle condition of* D, *and the local material properties*, $\varepsilon_{\mathrm{r}},\mu_{\mathrm{r}}$, *but independent of* $\boldsymbol{q}$, $h=\operatorname{diam}(\mathsf{D})$, *and* n, *such that*

$$\|\mathcal{P}_N\boldsymbol{T}^q\|_{\mathsf{D}}\leq C\frac{h^{\sigma-1}}{n^{p-2}}\|\boldsymbol{q}\|_{H^p(\mathsf{D})},$$

where $\sigma=\min(p,n+1)$.

Proof. We need only establish the result for $\mathcal{P}_N\boldsymbol{T}^E$, Eq. (24), as the result for $\mathcal{P}_N\boldsymbol{T}^H$ follows from identical arguments.

As $\mathcal{P}_N \boldsymbol{T}^E \in \mathsf{P}_n^3 = \sum_j \boldsymbol{T}_j^E L_j(x)$ we can multiply from the left with $\boldsymbol{T}_j^E$ and sum over all the nodes to recover

$$\|\mathcal{P}_N \boldsymbol{T}^E\|_{\mathsf{D}}^2 = (\mathcal{P}_N \boldsymbol{T}^E, \mathcal{P}_N \nabla \times (\boldsymbol{H} - \mathcal{P}_N \boldsymbol{H}))_{\mathsf{D}} - \frac{1}{\bar{Z}} (\mathcal{P}_N \boldsymbol{T}^E, \hat{\boldsymbol{n}} \times (Z^+ [\mathcal{P}_N \boldsymbol{H}] - \hat{\boldsymbol{n}} \times [\mathcal{P}_N \boldsymbol{E}]))_{\delta \mathsf{D}}.$$

Using the Cauchy–Schwarz inequality and the estimate [25]

$$\|\boldsymbol{q}_N\|_{\delta \mathsf{D}} \leq C \frac{n}{h^{1/2}} \|\boldsymbol{q}_N\|_{\mathsf{D}},$$

for all $\boldsymbol{q}_N \in \mathsf{P}_n^3(\mathsf{D})$, $h = \operatorname{diam}(\mathsf{D})$, we recover

$$\|\mathcal{P}_N \boldsymbol{T}^E\|_{\mathsf{D}} \leq C_1 \|\mathcal{P}_N \nabla \times (\boldsymbol{H} - \mathcal{P}_N \boldsymbol{H})\|_{\mathsf{D}} + C_2 \frac{n}{h^{1/2}} \frac{1}{\bar{Z}} \|Z^+ [\mathcal{P}_N \boldsymbol{H}_\tau] - [\mathcal{P}_N \boldsymbol{E}_\tau]\|_{\delta \mathsf{D}}, \quad (26)$$

where we for simplicity have introduced the tangential components

$$\boldsymbol{E}_\tau = \hat{\boldsymbol{n}} \times \boldsymbol{E}, \quad \boldsymbol{H}_\tau = \hat{\boldsymbol{n}} \times \boldsymbol{H}.$$

To bound the first term we use Lemma 4.1 to obtain

$$\|\mathcal{P}_N \nabla \times (\boldsymbol{H} - \mathcal{P}_N \boldsymbol{H})\|_{\mathsf{D}} \leq \|\nabla \times (\boldsymbol{H} - \mathcal{P}_N \boldsymbol{H})\|_{\mathsf{D}} \leq C \frac{h^{\sigma-1}}{n^{p-2}} \|\boldsymbol{H}\|_{H^p(\mathsf{D})}. \quad (27)$$

Consider now terms of the type

$$\|[\mathcal{P}_N \boldsymbol{E}_\tau]\|_{\delta \mathsf{D}} \leq \|\mathcal{P}_N \boldsymbol{E}_\tau^+ - \boldsymbol{E}_\tau^+\|_{\delta \mathsf{D}} + \|\mathcal{P}_N \boldsymbol{E}_\tau^- - \boldsymbol{E}_\tau^-\|_{\delta \mathsf{D}},$$

where $\boldsymbol{E}_\tau^+ = \boldsymbol{E}_\tau^- = \boldsymbol{E}_\tau$ represents the exact solution at $\delta \mathsf{D}$. Recalling the trace inequality [37]

$$\|q\|_{\delta \mathsf{D}}^2 \leq C \big(\|q\|_{\mathsf{D}} \|\nabla q\|_{\mathsf{D}} + h^{-1} \|q\|_{\mathsf{D}}^2 \big), \quad q \in H^1(\mathsf{D}),$$

implies that

$$\|q - \mathcal{P}_N q\|_{\delta \mathsf{D}}^2 \leq C \big(\|q - \mathcal{P}_N q\|_{\mathsf{D}} \|q - \mathcal{P}_N q\|_{H^1(\mathsf{D})} + h^{-1} \|q - \mathcal{P}_N q\|_{\mathsf{D}}^2 \big),$$

and we recover by combination with Lemma 4.1 the bound

$$\|[\mathcal{P}_N \boldsymbol{E}_\tau]\|_{\delta \mathsf{D}} \leq C \frac{h^{\sigma-1/2}}{n^{p-1}} \|\boldsymbol{E}\|_{H^p(\mathsf{D})}.$$

Combining this with Eqs. (26) and (27) one obtains the result

$$\|\mathcal{P}_N \boldsymbol{T}^E\|_{\mathsf{D}} \leq C_1 \frac{h^{\sigma-1}}{n^{p-2}} \|\boldsymbol{H}\|_{H^p(\mathsf{D})} + C_2 \frac{h^{\sigma-1}}{n^{p-2}} \big(\|\boldsymbol{E}\|_{H^p(\mathsf{D})} + \|\boldsymbol{H}\|_{H^p(\mathsf{D})} \big),$$

where (C_1, C_2) are independent of h and n but C_2 depends on the local material properties $(Z^{\pm}, Y^{\pm})$. ■

Let us now return to the original quest for an improved convergence estimate and consider the error equation

$$\left(L_i, \varepsilon\frac{\partial}{\partial t}(\mathcal{P}_N\boldsymbol{E}-\boldsymbol{E}_N)\right)_{\mathsf{D}} = (L_i, \mathcal{P}_N\nabla\times(\mathcal{P}_N\boldsymbol{E}-\boldsymbol{E}_N))_{\mathsf{D}} + \frac{1}{\bar{Z}}(L_i, \hat{\boldsymbol{n}}\times(Z^{+}[\mathcal{P}_N\boldsymbol{H}-\boldsymbol{H}_N] - \hat{\boldsymbol{n}}\times[\mathcal{P}_N\boldsymbol{E}-\boldsymbol{E}_N]))_{\delta\mathsf{D}} + (L_i, \mathcal{P}_N\boldsymbol{T}^{E})_{\mathsf{D}}, \tag{28}$$

for the electric field and similarly for the magnetic field

$$\left(L_i, \mu\frac{\partial}{\partial t}(\mathcal{P}_N\boldsymbol{H}-\boldsymbol{H}_N)\right)_{\mathsf{D}} = -(L_i, \mathcal{P}_N\nabla\times(\mathcal{P}_N\boldsymbol{H}-\boldsymbol{H}_N))_{\mathsf{D}} - \frac{1}{\bar{Y}}(L_i, \hat{\boldsymbol{n}}\times(Y^{+}[\mathcal{P}_N\boldsymbol{E}-\boldsymbol{E}_N] + \hat{\boldsymbol{n}}\times[\mathcal{P}_N\boldsymbol{H}-\boldsymbol{H}_N]))_{\delta\mathsf{D}} + (L_i, \mathcal{P}_N\boldsymbol{T}^{H})_{\mathsf{D}}. \tag{29}$$

The combination of these expressions with Lemma 4.4 and the methodology of the stability proof in Section 4.3 yields the improved convergence result

THEOREM 4.2. *Assume that a solution, $\boldsymbol{q}\in H^p(\mathsf{D}), p\geq 2$ to Maxwell's equations in $\Omega=\bigcup_k \mathsf{D}^k$ exists. Then the numerical solution, $\boldsymbol{q}_N$, to the semidiscrete approximation, Eqs.* (21)–(22), *converges to the exact solution, and the global error, $\sum_k \|\boldsymbol{q}-\boldsymbol{q}_N\|_{\mathsf{D}^k}$, is bounded as*

$$\begin{aligned}\sum_k \|\boldsymbol{q}(t)-\boldsymbol{q}_N(t)\|_{\mathsf{D}^k} &\leq C\sum_k\Big(\|\boldsymbol{q}(t)-\mathcal{P}_N\boldsymbol{q}(t)\|_{\mathsf{D}^k} \\ &\quad + \|\mathcal{P}_N\boldsymbol{q}(0)-\boldsymbol{q}_N(0)\|_{\mathsf{D}^k} + t\max_{s\in[0,t]}\|\boldsymbol{T}^{\boldsymbol{q}}(s)\|_{\mathsf{D}^k}\Big) \\ &\leq C\sum_k\left(\frac{h^{\sigma}}{n^{p}}\|\boldsymbol{q}(0)\|_{H^p(\mathsf{D}^k)} + t\frac{h^{\sigma-1}}{n^{p-2}}\max_{s\in[0,t]}\|\boldsymbol{q}(s)\|_{H^p(\mathsf{D}^k)}\right),\end{aligned}$$

where C depends on the material properties and the angle condition of the elements but not on h and n.

Proof. Since $\mathcal{P}_N\boldsymbol{E}-\boldsymbol{E}_N\in\mathsf{P}_n^3$ and $\mathcal{P}_N\boldsymbol{H}-\boldsymbol{H}_N\in\mathsf{P}_n^3$ we can use these as elementwise test functions in Eq. (28) and Eq. (29), respectively, and add these to obtain

$$\begin{aligned}&\frac{1}{2}\frac{d}{dt}((\mathcal{P}_N\boldsymbol{E}-\boldsymbol{E}_N, \varepsilon(\mathcal{P}_N\boldsymbol{E}-\boldsymbol{E}_N))_{\mathsf{D}} + (\mathcal{P}_N\boldsymbol{H}-\boldsymbol{H}_N, \varepsilon(\mathcal{P}_N\boldsymbol{H}-\boldsymbol{H}_N))_{\mathsf{D}}) \\ &\quad = \oint_{\delta\mathsf{D}}\Big(\hat{\boldsymbol{n}}\cdot(\mathcal{P}_N\boldsymbol{H}-\boldsymbol{H}_N)\times(\mathcal{P}_N\boldsymbol{E}-\boldsymbol{E}_N) + \frac{1}{\bar{Z}}(\mathcal{P}_N\boldsymbol{E}-\boldsymbol{E}_N)\cdot\hat{\boldsymbol{n}}\times(Z^{+}[\mathcal{P}_N\boldsymbol{H}-\boldsymbol{H}_N] \\ &\qquad -\hat{\boldsymbol{n}}\times[\mathcal{P}_N\boldsymbol{E}-\boldsymbol{E}_N]) - \frac{1}{\bar{Y}}(\mathcal{P}_N\boldsymbol{H}-\boldsymbol{H}_N)\cdot\hat{\boldsymbol{n}}\times(Y^{+}[\mathcal{P}_N\boldsymbol{E}-\boldsymbol{E}_N] \\ &\qquad +\hat{\boldsymbol{n}}\times[\mathcal{P}_N\boldsymbol{H}-\boldsymbol{H}_N])\Big)\,d\boldsymbol{x} + (\mathcal{P}_N\boldsymbol{E}-\boldsymbol{E}_N, \boldsymbol{T}^{E})_{\mathsf{D}} + (\mathcal{P}_N\boldsymbol{H}-\boldsymbol{H}_N, \boldsymbol{T}^{H})_{\mathsf{D}},\end{aligned}$$

where we have employed integration by parts once. Following the approach of Lemma 4.2 we sum over all the faces to obtain

$$\begin{aligned}\frac{1}{2}\frac{d}{dt}\sum_k((\mathcal{P}_N\boldsymbol{E}-\boldsymbol{E}_N,\varepsilon(\mathcal{P}_N\boldsymbol{E}-\boldsymbol{E}_N))_{\mathsf{D}^k}+(\mathcal{P}_N\boldsymbol{H}-\boldsymbol{H}_N,\varepsilon(\mathcal{P}_N\boldsymbol{H}-\boldsymbol{H}_N))_{\mathsf{D}^k})\\ \leq -\sum_k\left[\|[\mathcal{P}_N\boldsymbol{E}-\boldsymbol{E}_N]\|^2_{\mathsf{D}^k}+\|[\mathcal{P}_N\boldsymbol{H}-\boldsymbol{H}_N]\|^2_{\mathsf{D}^k}\right]\\ +\sum_k\left[(\mathcal{P}_N\boldsymbol{E}-\boldsymbol{E}_N,\mathcal{P}_N\boldsymbol{T}^E)_{\mathsf{D}^k}+(\mathcal{P}_N\boldsymbol{H}-\boldsymbol{H}_N,\mathcal{P}_N\boldsymbol{T}^H)_{\mathsf{D}^k}\right].\end{aligned}$$

Note that since ε and μ are uniformly bounded away from zero the material-weighted energy norm is L^2-equivalent. Furthermore, the term associated with the jump at the element interfaces is strictly negative and we recover the bound on the error

$$\frac{1}{2}\frac{d}{dt}\sum_k\|\mathcal{P}_N\boldsymbol{q}-\boldsymbol{q}_N\|^2_{\mathsf{D}^k}\leq C\sum_k(\mathcal{P}_N\boldsymbol{q}-\boldsymbol{q}_N,\mathcal{P}_N\boldsymbol{T}^q)_{\mathsf{D}^k},$$

which, by using the Cauchy–Schwarz inequality and integration in time, yields the result

$$\sum_k\|\mathcal{P}_N\boldsymbol{q}(t)-\boldsymbol{q}_N(t)\|_{\mathsf{D}^k}\leq C\sum_k\left(\|\mathcal{P}_N\boldsymbol{q}(0)-\boldsymbol{q}_N(0)\|_{\mathsf{D}^k}+t\max_{s\in[0,t]}\|\mathcal{P}_N\boldsymbol{T}^{\boldsymbol{q}}(s)\|_{\mathsf{D}^k}\right).$$

Now combining this with Lemma 4.3 and Lemma 4.4 establishes the result and proves convergence on weak assumptions of local, elementwise smoothness of the solution. ■

We have hence established the semidiscrete result that the error cannot grow faster than linearly in time and that we can control the growth rate by increasing the resolution provided the solution is sufficiently regular. As we shall verify in Section 5 this linear growth is a sharp result.

Prior to that, a few comments are in place. A rigorous generalization of the results obtained above to include situations with general curvilinear elements and/or spatial variation of the materials within each element is not straightforward. This is due to the generation of higher-order polynomials from the products of the individual polynomial expressions of the fields, the materials, and the geometry. One can, however, gain an intuitive understanding of how the geometry and the material variations may impact the accuracy by assuming that the polynomial representations are not of the fields only but rather of the combined functions, $\sqrt{J}(\sqrt{\varepsilon_r}\boldsymbol{E},\sqrt{\mu_r}\boldsymbol{H})$. In this setting, we are working only with n-order polynomial expansions and one can expect that the overall picture from the results derived above will hold also for these new functions. Hence, where we originally had an nth order polynomial to represent the fields, $(\boldsymbol{E},\boldsymbol{H})$, we are now left with an nth order polynomial to represent the combined variation. One consequence of this is that we loose accuracy when considering only the fields as we essentially must share the resolution power between the fields, the geometry, and the material variation. In particular, if the element is strongly distorted, i.e., J varies significantly, one can expect loss of accuracy as compared to the straightsided approximation. Provided, however, that the geometry is smooth, i.e., J nonsingular, and the local material variation is smooth, spectral convergence is preserved.

4.5. *Convergence of Divergence Error*

In the absence of sources, it is well known that the electric and the magnetic fields must remain solenoidal throughout the computation. An assumption to this effect was indeed imposed by choosing to only solve the time-dependent part of Maxwell's equations, Eq. (3), and leaving the divergence conditions as consistency conditions on the initial conditions. However, given that we cannot expect to recover the projection of the analytic solution but rather will compute a different, albeit convergent, solution we need to consider the divergence of this numerical solution to justify this.

Using the results of Section 4.4 we can state:

THEOREM 4.3. *Assume that a solution, $\boldsymbol{q} \in H^p(\mathsf{D}), p \geq 4$ to Maxwell's equations in $\Omega = \bigcup_k \mathsf{D}^k$ exists. Then there exists a constant, C, dependent on p and the angle condition of D^k, but independent of $\boldsymbol{q}$, $h = \operatorname{diam}(\mathsf{D})$, and n, such that the divergence of the numerical solution, $\boldsymbol{q}_N$, to the semidiscrete approximation, Eqs.* (21)–(22), *is bounded as*

$$\sum_k \|\nabla \cdot \boldsymbol{q}_N(t)\|_{\mathsf{D}^k} \leq C \sum_k \left(\frac{h^{\sigma-1}}{n^{p-2}} \|\boldsymbol{q}(0)\|_{H^p(\mathsf{D}^k)} + t \frac{h^{\sigma-2}}{n^{p-4}} \max_{s\in[0,t]} \|\boldsymbol{q}(s)\|_{H^p(\mathsf{D}^k)} \right),$$

where $\sigma = \min(p, n+1)$.

Proof. The result follows directly from Theorem 4.2 and the inverse inequality [25]

$$\|\nabla \cdot u_N\|_{\mathsf{D}} \leq \frac{n^2}{h} \|u_N\|_{\mathsf{D}},$$

for all $u_N \in P_n^3(\mathsf{D})$. ■

As could be expected, the result inherits the temporal linear growth from the convergence result and confirms the possibility of recovering spectral convergence of the divergence under the assumption of sufficient smoothness of the solutions. It should be noted that while the result confirms high-order accuracy and convergence, the estimate for the actual convergence rate may well be suboptimal and certainly require more regularity than can be guaranteed.

4.6. *Entràcte on the Scattered Field Formulation*

Let us briefly return to an analysis of the scattered field formulation discussed in Section 2.1, with the modified scattered field equations given in Eqs. (6) and (7). We recall that we split the solution, $\boldsymbol{q}$, as

$$\boldsymbol{q} = \boldsymbol{q}^{\mathrm{s}} + \boldsymbol{q}^{\mathrm{i}},$$

and exploit the linearity of Maxwell's equations to solve for the scattered field, $\boldsymbol{q}^{\mathrm{s}}$, subject to the forcing by the incident field, $\boldsymbol{q}^{\mathrm{i}}$. As discussed in Section 2.1, this does not alter the scheme in any significant way except at metallic boundaries where the boundary condition on the electric field component takes the form

$$\boldsymbol{n} \times \boldsymbol{E}_N^{\mathrm{s},+} = -\boldsymbol{n} \times \boldsymbol{E}_N^{\mathrm{s},-} - 2\mathcal{P}_N \boldsymbol{E}^{\mathrm{i}},$$

in the notation of Lemma 4.2, while the boundary condition on the magnetic field remains

$$\boldsymbol{n} \times \boldsymbol{H}_N^{\mathrm{s},+} = \boldsymbol{n} \times \boldsymbol{H}_N^{\mathrm{s},-}.$$

Since this constitutes the only difference, we can restrict the subsequent analysis to the case of a metallic object in vacuum without loss of generality as all other complications are covered by the analysis of the total field scheme.

THEOREM 4.4. *Assume that a scattered field solution, $\boldsymbol{q}^{\mathrm{s}} \in H^p(\mathsf{D}), p \geq 2$ to Maxwell's equations in $\Omega = \bigcup_k \mathsf{D}^k$ exists, and that the incident field $\boldsymbol{q}^{\mathrm{i}} \in H^p(\mathsf{D}), p \geq 2$. Then the energy of the numerical scattered field solution, $\boldsymbol{q}_N^{\mathrm{s}}$, to the semidiscrete approximation of Eqs.* (6) *and* (7) *is bounded as*

$$\sum_k \left\|\boldsymbol{q}_N^{\mathrm{s}}(t)\right\|_{\mathsf{D}^k} \leq C \sum_k \left(\|\mathcal{P}_N \boldsymbol{q}^{\mathrm{i}}(t)\|_{\mathsf{D}^k} + \left\|\mathcal{P}_N \boldsymbol{q}^{\mathrm{i}}(0) + \boldsymbol{q}_N^{\mathrm{s}}(0)\right\|_{\mathsf{D}^k} + t \max_{s\in[0,t]} \left\|\boldsymbol{T}^{\boldsymbol{q}^{\mathrm{i}}}(s)\right\|_{\mathsf{D}^k} \right),$$

where C depends on the material properties and the angle conditions of the elements but not on h and n.

Proof. The proof proceeds in a way very similar to that of Theorem 4.2 and is left out. ■

Hence, also the scattered fields remain bounded up to linear growth in time. An interesting difference between this result on that of Theorem 4.2 for the total field formulation is that the accuracy and growth rate of the former is controlled solely by the smoothness of the incident field with the potential for exponential convergence for sufficiently smooth illuminating fields.

5. VALIDATION AND PERFORMANCE OF THE SCHEME

Having developed the formulation for the time-domain solution of Maxwell equations it is now time to consider the actual performance of the computational framework. Unless stated otherwise, all computational are done in the scattered fields formulation.

In the following we shall discuss the validity of the main theoretical results through a few examples as well as exemplify the versatility and overall accuracy and performance of the complete framework for a number of benchmarks. Temporal integration of the semidiscrete approximation given in Eqs. (18) and (19) is done using a fourth-order, five-stage low-storage Runge–Kutta scheme [38] and a stability-limited time-step scaling as

$$\Delta t \leq \mathrm{CFL} \min_{\Omega} \sqrt{\varepsilon_{\mathrm{r}}\mu_{\mathrm{r}}} |\chi|^{-1},$$

with $\sqrt{\varepsilon_{\mathrm{r}}\mu_{\mathrm{r}}}$ reflecting the modified local speed of light due to materials and

$$\chi = \frac{|\nabla\xi|}{\Delta\xi} + \frac{|\nabla\eta|}{\Delta\eta} + \frac{|\nabla\zeta|}{\Delta\zeta}.$$

Here $|\cdot|$ refers to the absolute value of each and of the vector components, i.e., $|\nabla\xi| = [|\xi_x|, |\xi_y|, |\xi_z|]^{\mathrm{T}}$. Hence, χ provides a measure of the local grid distortion as a consequence of the mapping, Ψ, of I into D, and $(\Delta\xi, \Delta\eta, \Delta\zeta)$ measures the axial distance separating neighboring nodal points in I. In this setting CFL typically takes values of $\mathcal{O}(1)$ while the

time step, Δt, scales as $\Delta t \simeq l/n^2$, where l is the minimum edge length on all tetrahedra and n is the polynomial order of the approximation.

As a general measure of error we shall use the discrete L^p-norm of the error defined as

$$\|\delta f(t)\|_p = \left(\sum_{j,k} \left[f_N(\boldsymbol{x}_j^k, t) - f(\boldsymbol{x}_j^k, t) \right]^p \right)^{1/p},$$

where $f_N(\boldsymbol{x}, t)$ is the numerical approximation to the exact value, $f(\boldsymbol{x}, t)$, summed over all nodes, j, within each of the k elements.

5.1. *Elementary Tests and Verification of Theoretical Results*

As a first verification of the theoretical estimates, and in particular the linear growth predicted in Theorem 4.2, we consider the solution of the two-dimensional Maxwell's equations in the TM-polarization; i.e., we solve for (H_x, H_y, E_z). There is, however, nothing special about this polarization.

The computational problem is that of a simple two-dimensional vacuum-filled cavity, assumed to be defined by $(x, y) \in [-1, 1] \times [-0.25, 0.25]$, with the walls at $x = \pm 1$ taken to be perfect electrical conductors while the cavity is assumed to be periodic in the y-direction. The initial condition is a simple oscillatory cavity solution as

$$H_x(x, y, 0) = 0, \quad H_y(x, y, 0) = \cos(\pi x), \quad E_z(x, y, 0) = 0,$$

and the computational domain is discretized by eight equivalent isosceles, each with 0.5 wavelength long sides.

In Fig. 2 we show the temporal envelope of the maximum error of $H_y(t)$, computed using the same eight elements while increasing the order of the approximation. Following the main result, Theorem 4.2, we expect that the error can grow at most linearly in time and that the growth rate should vanish spectrally for a smooth solution. The results in Fig. 2 not only confirm the validity of both statements but also illustrate that Theorem 4.2 is sharp; i.e., we

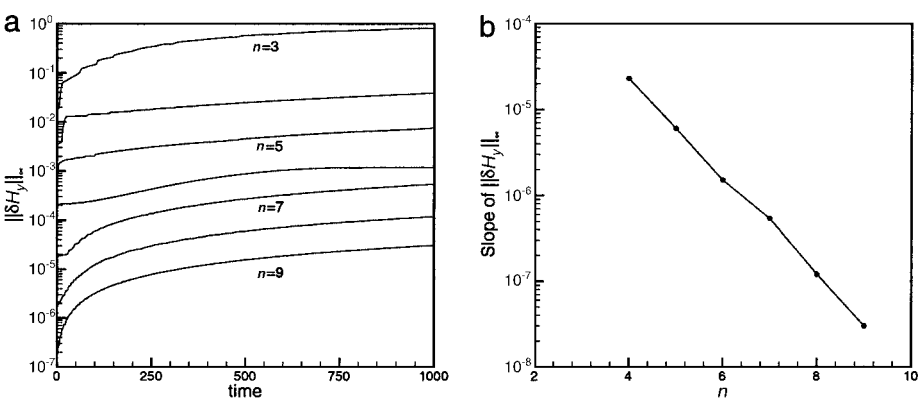

FIG. 2. In (a) is shown the temporal envelope of the maximum error on $H_y(t)$ in the two-dimensional cavity for different orders, n, of the approximation. The slope of the linear growth is plotted in (b), confirming spectral convergence as predicted in Theorem 4.2.

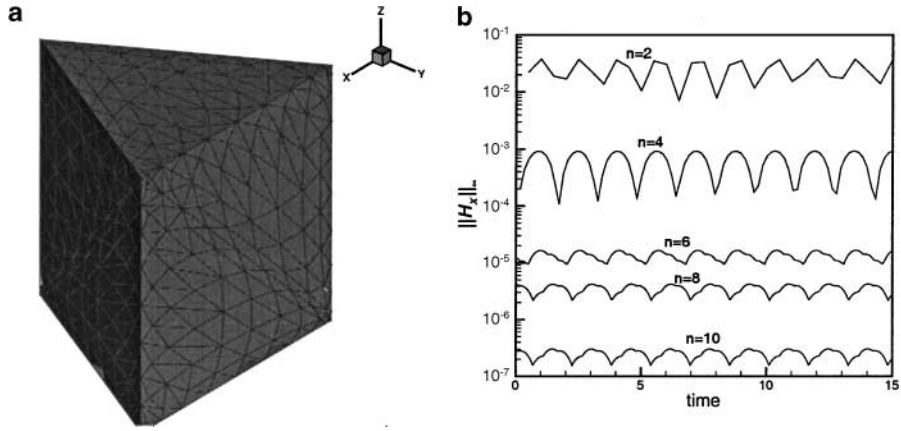

FIG. 3. In (a) we illustrate the prism tiled using three high-order tetrahedra while (b) illustrates the maximum of H_x for a (y, z)-polarized plane wave propagation as a function of time and order of the approximation, n, confirming spectral convergence for the three-dimensional case.

cannot in general guarantee slower than linear error growth, although we can control the growth rate by the order of the approximation.

The ability to propagate waves over very long distances is likewise confirmed in Fig. 2. Using as guideline that two edges span a wavelength, we see that with 7 points per wavelength (two $n = 3$ triangles) yields $\mathcal{O}(1)$ error after 500 periods. Only 9 points per wavelength (two $n = 4$ triangles) results in about 1% error while 11 points per wavelength (two $n = 5$ triangles) ensures about 0.1% error after 500 periods. This is a testament to the advantage of using a high-order framework for wave propagation problems.

Let us also consider a simple three-dimensional test case in which we have tiled a straightfaced prism using three straightfaced tetrahedra as illustrated in Fig. 3. The test is that of a plane wave propagating through the prism with the exact solution being used as the boundary conditions. As shown in Fig. 3 we recover a rapid exponential convergence as the order, n, of the approximation is increased.

5.2. *Two-Dimensional Examples*

Having verified the performance of the basic computational setup as well as the theoretical estimates, let us now consider problems of a less simple character. This shall not only allow us to illustrate more general features of the proposed framework but shall also be used to verify that all the properties of the high-order unstructured grid formulation, seen so convincingly in the last section for simple examples, carry over to the solution of more realistic problems.

We shall focus attention on problems described by the two-dimensional TM-polarized Maxwell's equations in the form

$$\begin{aligned} \mu_r \frac{\partial H_x}{\partial t} &= -\frac{\partial E_z}{\partial y}, \\ \mu_r \frac{\partial H_y}{\partial t} &= \frac{\partial E_z}{\partial x}, \\ \varepsilon_r \frac{\partial E_z}{\partial t} &= \frac{\partial H_y}{\partial x} - \frac{\partial H_x}{\partial y}, \end{aligned} \tag{30}$$

subject to boundary conditions between two regions with material parameters, $\varepsilon_r^{(k)}$ and $\mu_r^{(k)}$, for $k = 1, 2$, as

$$\hat{\boldsymbol{n}} \times \boldsymbol{H}^{(1)} = \hat{\boldsymbol{n}} \times \boldsymbol{H}^{(2)},$$
$$E_z^{(1)} = E_z^{(2)}.$$

Here $\boldsymbol{H}^{(k)} = (H_x^{(k)}, H_y^{(k)}, 0)^{\mathrm{T}}$, and $\hat{\boldsymbol{n}} = (\hat{n}_x, \hat{n}_y, 0)^{\mathrm{T}}$ represents a unit vector normal to the interface. For the case of a perfectly conducting metallic boundary the condition becomes particularly simple as

$$E_z = 0.$$

The computational domain is truncated with a Cartesian PML [39] using a quadratic absorption profile.

It is worthwhile emphasizing that results of equal quality and accuracy as those shown in the following for the TM-polarized case have been obtained for the TE-polarized case.

As a first example we consider that of plane wave scattering by a perfectly conducting circular cylinder with a radius of $a = 7.5\lambda$, i.e., $ka = 15\pi$. The surrounding medium is assumed to be vacuum, i.e., $\varepsilon_r = \mu_r = 1$. The finite element grid, consisting of 854 triangles, used for this computation is shown in Fig. 4 along with a section of the grid illustrating the body-conforming nature of the approximation as well as the nodal grid supporting the high-order approximation. Prony extrapolation [40] is used to reduce the required computing time to reach the harmonic steady state.

In Fig. 5 we compare the computed bistatic radar cross section, RCS(θ), with the exact series solution [41], for various orders, n, of the approximation using the finite element grid in Fig. 4. As expected we observe a very rapid convergence with increasing n, yielding a reasonable engineering accuracy computation with the fourth-order scheme while increasing the order to $n = 8$ results in a perfect match. A quantitative confirmation of this is also shown in Fig. 5, illustrating exponential convergence of the RCS with increasing n.

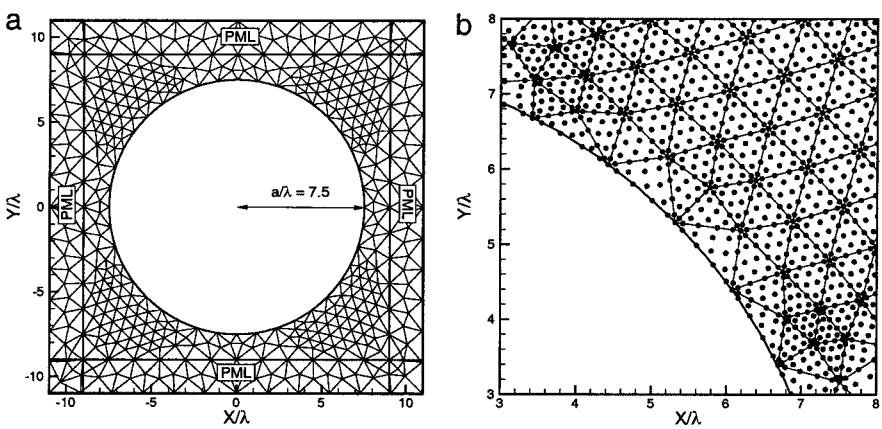

FIG. 4. In (a) is shown the finite element grid, consisting of 854 triangles, used for computing scattering by a perfect electrically conducting cylinder of size $ka = 15\pi$. A section of the grid in (b) illustrates the body-conforming nature of the grid and the nodal grid supporting the high-order approximation.

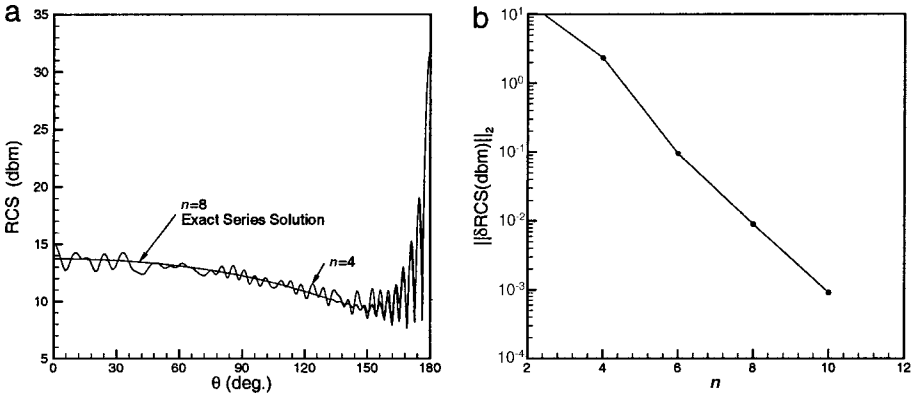

FIG. 5. In (a) is shown the bistatic radar cross section, RCS(θ), as computed using the exact series representation as well as the unstructured grid method at different polynomial orders, n. Evidence of high-order convergence for the RCS computation is given in (b), showing exponential decay of the error in RCS (dBm) with increasing order of the approximation.

One of the most appealing advantages of a high-order framework on simplices is the ability to import a strongly skewed and/or coarse finite element grid and recover a fully converged solution by increasing the order of the approximation rather than having to reconstruct an improved finite element discretization. This property is particularly important and useful for large three-dimensional problems where the grid generation phase often is complex and time consuming. As an illustration of this approach to convergence, we consider in Fig. 6 plane wave scattering by a PEC cylinder with a radius of one wavelength, i.e., $ka = 2\pi$. A measure of accuracy is based on the observation that the symmetry of the problem makes one expect the scattered fields themselves to maintain a high degree of symmetry.

This is indeed confirmed in Fig. 6 where we show a deliberately chosen coarse and skewed grid and the rapid recovery of the symmetry of one of the scattered field components, H_x, as the order, n, of the approximation is increased without modifying the underlying finite element grid. The detail to which the symmetry is restored is particularly noteworthy.

As an illustration of the capability to handle materials let us consider plane wave scattering by a penetrable circular cylinder with a radius of $a = 3.5\lambda$ consisting of an ideal dielectric with $\varepsilon_r = 2.0$, i.e., similar to that of glass. The problem is again solved in a pure scattered field formulation. The fully body-conforming finite element discretization, consisting of a total of 1020 triangles, is illustrated in Fig. 7. We note that the absorbing PML layer, containing about $\frac{2}{3}$ of the total amount of triangles, is unnecessarily thick for illustration only and can be decreased without loss of accuracy.

As is likewise illustrated in Fig. 7 we recover the full bistatic radar cross section, RCS(θ), with excellent correspondence to the exact solution [42] and quantitative agreement over a 40-db dynamic range.

5.3. *Three-Dimensional Examples*

As a first verification of the general three-dimensional framework, let us consider plane wave scattering by a $ka = 10$ perfectly conducting sphere, the analytic solution of which is given by a Mie series [41].

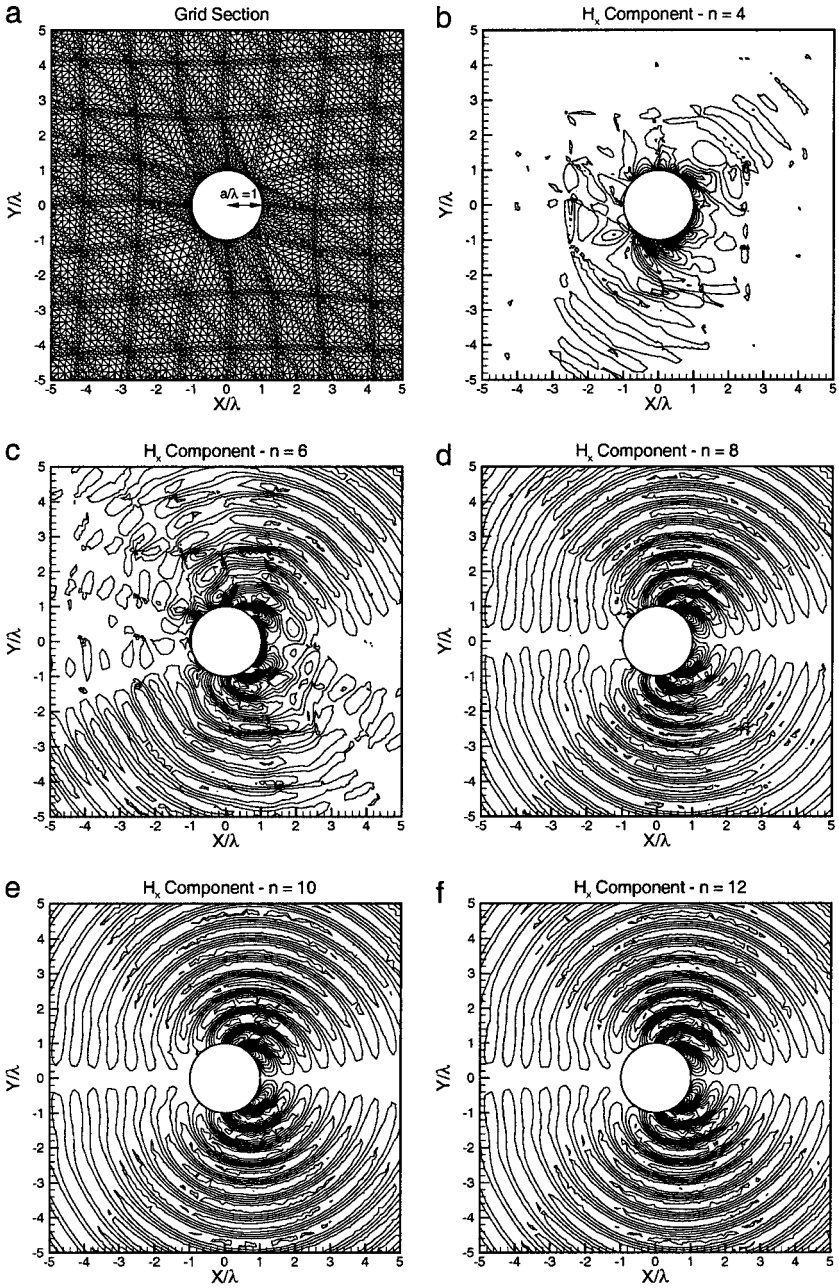

FIG. 6. An example of convergence by increasing the order of the approximation, n, on a deliberately chosen skewed finite element grid, illustrated in (a). The convergence is illustrated in (b)–(f) when increasing the order from fourth order to 12th order, showing a complete recovery of the expected symmetry of the scattered field component, H_x.

We use a fully bodyconforming grid with a total of 3000 elements, having an average edge length at the sphere of $4\lambda/5$. In contrast to the two-dimensional case where we used a PML to truncate the computational domain we choose in the three-dimensional case to embed the sphere in a $(20\lambda)^3$ cube and employ stretching of the elements toward the

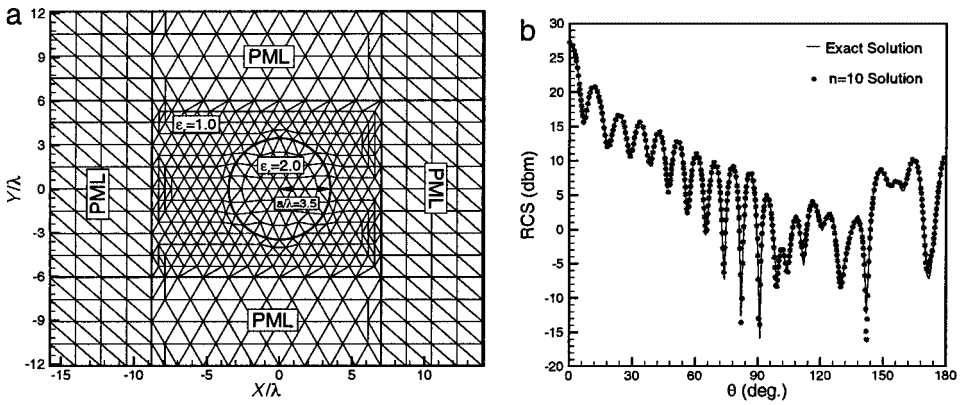

FIG. 7. Plane wave scattering by a $ka = 7\pi$ dielectric circular cylinder with a relative permittivity $\varepsilon_r = 2.0$. In (a) we show the finite element discretization while (b) shows a comparison between the computed bistatic radar cross section, RCS(θ), obtained with a 10th order approximation and that recovered by evaluating the exact solution.

outer boundary. The grid is stretched such that the average edge is about 2λ at the outer boundary. As in the two-dimensional case, all examples are done using a fourth-order low-storage Runge–Kutta scheme to advance in time and Prony extrapolation to identify the solution.

In Fig. 8 we illustrate the convergence of the scheme with a fixed grid when increasing the order of the approximation within each tetrahedron. Even for $n = 3$, i.e., a third-order scheme with about 5 points per wavelength, we compute a reasonable solution while increasing the order yields a rapidly converging solution as one would expect.

As a considerably more challenging problem, let us consider scattering by a perfectly conducting business card-sized metallic plate as illustrated in Fig. 9. The horizontally polarized plane wave impinges at the metallic plate at an almost grazing angle, causing the excitation of strong waves along the edges of the metallic plate as well as along the length of the plate. These waves contribute significantly to the scattering process and need to be resolved to accurately predict the far-field scattering.

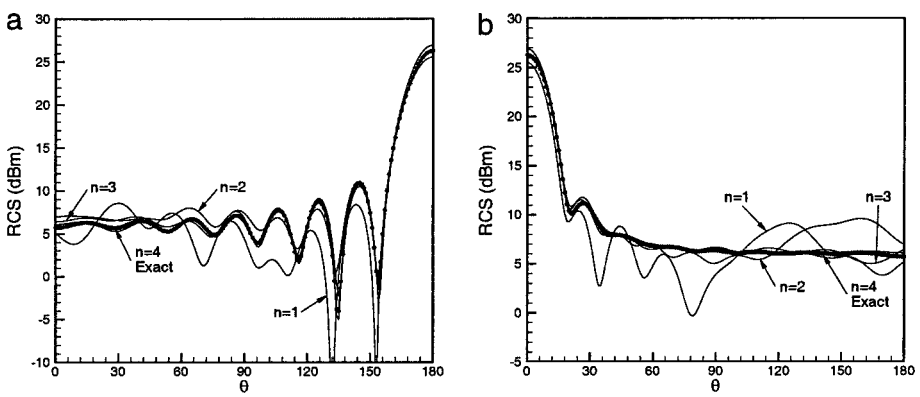

FIG. 8. Plane wave scattering by a $ka = 10$ metallic sphere for a fixed grid and increasing order, n, of the polynomial approximation. In (a) we show the convergence of RCS(θ, 0) for vertical polarization (TM), while (b) shows RCS(θ, 90) for horizontal polarization (TE) of the incident field.

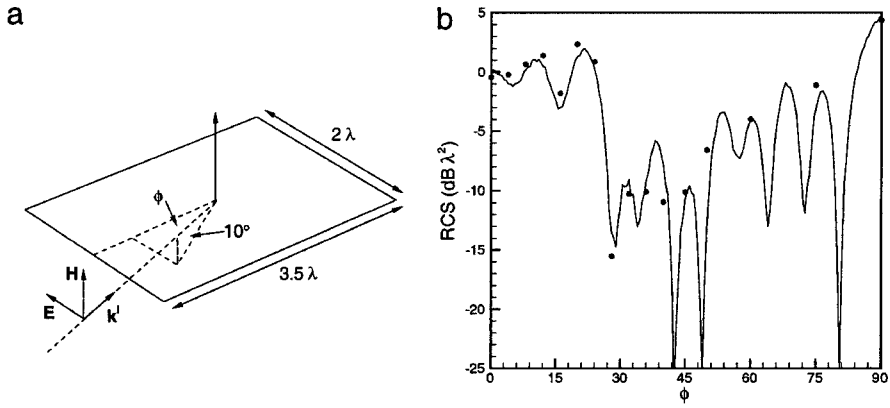

FIG. 9. In (a) we show the geometry for the plane wave scattering by a metallic business card while (b) shows the comparison between monostatic RCS experimental results [43] (full line) for horizontal polarization of the illuminating field and computed data points (·).

This problem, being one of the EMCC benchmark problems [43] for code validation, is addressed by using a total of 27,000 straightsided tetrahedra, each supporting a fourth-order polynomial approximation. The average edge length at the edge of the business card is approximately $\lambda/5$. The metallic plate is embedded in a $(20\lambda)^3$ cube, with the elements being stretched to about 4λ at the outer boundary.

In Fig. 9 we also show the comparison between the experimentally measured monostatic RCS [43] and a number of computed data points. Again we observe good agreement over the full azimuthal range with results well within the experimental error. We note in particular the good agreement in the backscatter region where the scattering is dominated by traveling waves.

As a final example we consider plane wave scattering by a dielectric cylinder of finite length. As illustrated in Fig. 10, the length of the cylinder is 5λ and the nonmagnetic material has a permittivity of $\varepsilon_r = 2.25$, similar to that of glass. The nature of the fields

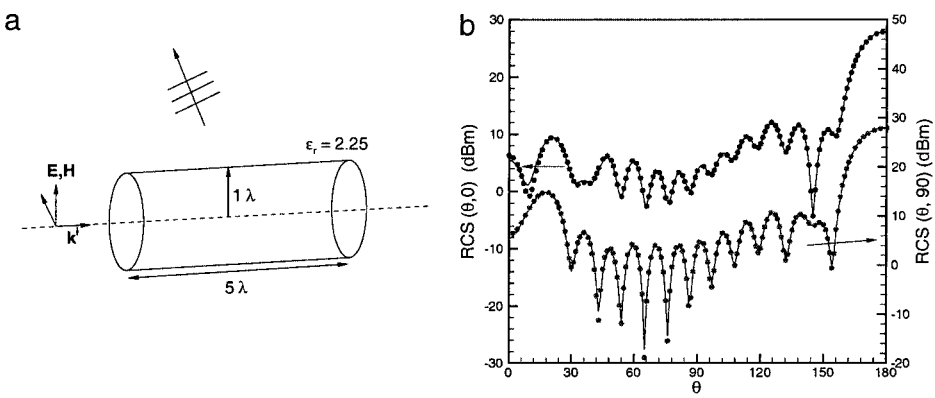

FIG. 10. In (a) we show the geometry for the plane wave scattering by a dielectric finite length cylinder while (b) shows the RCS $(\theta, 0)$ for vertical polarization (·) of the illuminating field and RCS $(\theta, 90)$ for horizontal polarization (·) compared with results obtained using a pseudospectral axisymmetric code (full line) [12].

is less dramatic than in the previous case, and we find that using a total of approximately 67,000 elements, supporting a fourth-order approximation and with an average vacuum edge length at the cylinder of $\lambda/3$, suffices to accurately predict the far-field scattering. The full computational domain is a cylinder of radius 16λ and length 23λ with the stretched elements having a average length of 4λ at the outer boundary.

In Fig. 10 we show a direct comparison between the full bistatic RCS for a plane wave impinging directly at the end of the cylinder as computed using the current framework and an independently verified pseudospectral multidomain axisymmetric code [12]. As expected we find an almost perfect agreement between the results of the two schemes over approximately a 50-dB dynamic range.

5.4. *Parallel Performance*

The discontinuous element formulation of the scheme enables a highly efficient implementation at contemporary large-scale distributed memory machines. While this is a lesser concern for two-dimensional problems, it is essential for enabling the modeling of large-scale three-dimensional problems.

The scheme is implemented in C with all computationally intensive parts based extensively on Level 3 BLAS [44]. The parallel interface is written in MPI [45] with METIS [46] to distribute the elements over the processors. To ensure high cache efficiency, we employ bandwidth minimization [47] of the nodal points locally to the processors [48]. For computations maximizing the capacity of the processors, i.e., filling the local memory, this is critical to ensure high performance.

In Table I we list the parallel speedup relative to the $n = 2$ case as the number of processors are increased. A few things are worth noting. For a fixed size problem, the parallel speedup decreases slightly as the number of processors increases as is natural since the relative communication cost increases. On the other hand, for problem sizes utilizing the available resources we find a high parallel efficiency; e.g., increasing the problem size and the number of processors yields a close to constant speedup. The data also shows a minor decrease in relative performance for high order on many processors, which we speculate is related to cache effects known to be become important as the size of the operators increase [29]. We generally observe better than 90% parallel efficiency, consistent with other similar studies [49].

TABLE I
Parallel Speedup for a 123.000 Element Grid, Scaled to Timing for $n = 2$ on 4 Processors

Polynomial order (n)	Degrees of freedom ($\times 10^6$)	Number of processors				
		4	8	16	32	64
2	7.4	1.0	2.0	3.9	7.5	13.7
3	14.8	—	0.9	1.8	3.5	6.4
4	25.8	—	—	1.0	1.9	3.6
5	41.3	—	—	—	0.8	1.6

Note. — Implies insufficient memory local to the nodes.

6. CONCLUDING REMARKS AND OUTLOOK

The main purpose of this paper has been to introduce the reader to a new class of high-order unstructured grid methods suitable for solving systems of linear conservation laws with special attention paid to the time-domain solution of Maxwell's equations. A number of central elements separate the current framework from previous attempts to develop high-order accurate methods on unstructured tetrahedral grids. The use of a purely nodal basis has a number of advantages in terms of ease of implementation by simple matrix–vector operations as well as the promise to yield a highly efficient implementation. Furthermore, the generalized discontinuous penalty scheme was introduced, offering an inherently parallel discontinuous formulation with a purely block-diagonal mass matrix that can be inverted during preprocessing.

The particular focus on Maxwell's equations allowed us to develop a complete, if not optimal, convergence theory. A similar analysis can be completed for other classes of linear problems such as acoustics and linear elasticity. We have confirmed the results of the analysis by thorough computational experiments, illustrating the flexibility, versatility, and efficiency of the proposed high-order accurate unstructured grid framework.

While we have focused on linear conservation laws, the central elements of the framework allow for more general formulations that enable the solution of typical systems of nonlinear conservation laws. This naturally raises questions about proper formulation of the fluxes at interfaces, conservation, and stability of high-order schemes when approximating problems with discontinuous solutions. We shall address these issues in [30] where we shall also demonstrate the performance of such generalized formulations for the solution of nonlinear conservation laws.

APPENDIX: EFFICIENT AND ACCURATE IMPLEMENTATION TECHNIQUES

From the discussions in Section 3.2 it is clear that the Vandermonde matrix, $\mathbf{V}$, plays a crucial role when setting up the discrete operators for interpolation and differentiation. The properties of $\mathbf{V}$, e.g., its conditioning, depends exclusively on the structure of nodal set, $\boldsymbol{\xi}_j$, and on the way in which we choose to represent the basis, i.e., $p_i(\boldsymbol{\xi})$. While the former is chosen to ensure well-behaved Lagrange interpolation polynomials, we have significant freedom in the specification of $p_i(\boldsymbol{\xi})$.

A particularly simple choice is that of the multivariate monomial basis, i.e., $p_i(\boldsymbol{\xi}) = \xi^i \eta^j \zeta^k$. However, even for interpolation in one dimension, i.e., $p_i(\xi) = \xi^i$, it is well known that this basis leads to the classical Vandermonde matrix, $\mathbf{V}$, with an exponentially growing condition number. Hence, even for moderate values of n we can expect severe problems when attempting to compute the action of $\mathbf{V}^{-1}$. The well-known solution to this problem is to choose a basis that is orthonormalized with respect to some proper inner product to assure the maximum degree of linear independence of the basis.

Such a basis has been known for long [50–52] and takes the form

$$\psi_{ijk}(\boldsymbol{\xi}) = P_i^{(0,0)}(r)\left(\frac{1-s}{2}\right)^i P_j^{(2i+1,0)}(s)\left(\frac{1-t}{2}\right)^{i+j} P_k^{(2i+2j+2,0)}(t), \tag{A.1}$$

where

$$r = -\frac{2(1+\xi)}{\eta+\zeta} - 1, \quad s = \frac{2(1+\eta)}{1-\zeta} - 1, \quad t = \zeta,$$

and $P_n^{(\alpha,\beta)}(x)$ signifies the classical Jacobi polynomial of order n [53].

The tensor product structure of the basis, Eq. (A.1), becomes evident when one realizes that while $\boldsymbol{\xi}$ is restricted by I, the mapped coordinates, (r, s, t), cover $[-1, 1]^3$. Furthermore, it is easy to see that the polynomial space, P_n^3, is

$$\mathsf{P}_n^3 = \text{ span}\{\psi_{ijk}(\boldsymbol{\xi}); i, j, k \geq 0; i + j + k \leq n\}.$$

An important property of the basis, Eq. (A.1), is its orthogonality on I [21] as

$$\int_{\mathsf{I}} \psi_{ijk}(\boldsymbol{\xi})\psi_{pqr}(\boldsymbol{\xi})\, d\boldsymbol{\xi} = \gamma_{ijk}\delta_{ijk,pqr},$$

where $\delta_{ijk,pqr}$ is the multidimensional Dirac delta and the normalization is

$$\gamma_{ijk} = \frac{2}{2i+1}\frac{2^{2i+2}}{2(i+j)+2}\frac{2^{2(i+j)+3}}{2(i+j+k)+3}.$$

Let us introduce the index, $\alpha \in [0, N]$, reflecting some chosen ordering of (i, j, k) and hence ψ_{ijk}. We can thus rename the polynomial basis $\psi_{ijk}(\boldsymbol{\xi}) = \psi_\alpha(\boldsymbol{\xi})$ to simplify the notation.

Using the orthogonal basis, ψ_α, it is natural to define the Vandermonde matrix to have the entries

$$\mathbf{V}_{ij} = \frac{1}{\sqrt{\gamma_j}}\psi_j(\boldsymbol{\xi}_i).$$

The relation between the nodal and the modal representation of a function, $\boldsymbol{f}$, follows directly from Eq. (11) as

$$\boldsymbol{f} = \mathbf{V}\hat{\boldsymbol{f}}, \quad \hat{\boldsymbol{f}} = \mathbf{V}^{-1}\boldsymbol{f}.$$

Furthermore, we can compute the entries of the stiffness matrices, Eq. (20), directly by defining the entries using the derivatives of $\psi_i(\boldsymbol{\xi})$ expressed by the identity [53]

$$\frac{d}{d\xi}P_n^{(\alpha,0)}(\xi) = \frac{1}{2}(n+1+\alpha)P_{n-1}^{(\alpha+1,1)}(\xi),$$

and that

$$\nabla \boldsymbol{L} = (\mathbf{V}^{-1})^{\mathrm{T}}\nabla\boldsymbol{\psi}.$$

We also need to evaluate inner products on the general curvilinear tetrahedron; i.e., we need an efficient and accurate procedure for computing

$$(f_N, g_N)_{\mathsf{D}} = \int_{\mathsf{I}} f_N(\boldsymbol{\xi})g_N(\boldsymbol{\xi})J(\boldsymbol{\xi})\, d\boldsymbol{\xi},$$

where J refers to the transformation Jacobian for the mapping between D and I and $f_N \in P_n^3$, $g_N \in P_n^3$. To evaluate this inner product, we exploit that f_N and g_N are expressed uniquely by their expansion in Lagrange polynomials as

$$(f_N, g_N)_{\mathsf{D}} = \sum_{i,j=0}^{N} f_i g_j \int_{\mathsf{I}} L_i(\boldsymbol{\xi}) L_j(\boldsymbol{\xi}) J(\boldsymbol{\xi})\, d\boldsymbol{\xi}.$$

Furthermore, using the basis itself, $\psi_\alpha(\boldsymbol{\xi})$, we can express the Lagrange polynomials themselves using Eq. (13) on the form

$$L_i(\boldsymbol{\xi}) = \sum_{k=0}^{N} \mathbf{V}_{ik}^{-1} \psi_k(\boldsymbol{\xi}).$$

This immediately yields the expression

$$\begin{aligned}(f_N, g_N)_{\mathsf{D}} &= \sum_{i,j=0}^{N} f_i g_j \sum_{k,l=0}^{N} \mathbf{V}_{ki}^{-1} \mathbf{V}_{lj}^{-1} \int_{\mathsf{I}} \psi_k(\boldsymbol{\xi}) \psi_l(\boldsymbol{\xi}) J(\boldsymbol{\xi})\, d\boldsymbol{\xi} \\ &= \sum_{i,j=0}^{N} f_i g_j \sum_{k,l=0}^{N} \mathbf{V}_{ki}^{-1} \mathbf{V}_{lj}^{-1} \mathbf{W}_{kl}, \end{aligned} \tag{A.2}$$

where the symmetric matrix of weights, $\mathbf{W}$, has the entries

$$\mathbf{W}_{kl} = \int_{\mathsf{I}} \psi_k(\boldsymbol{\xi}) \psi_l(\boldsymbol{\xi}) J(\boldsymbol{\xi})\, d\boldsymbol{\xi}.$$

On matrix form Eq. (A.2) becomes

$$(f_N, g_N)_{\mathsf{D}} = (\mathbf{V}^{-1}\boldsymbol{f})^{\mathrm{T}} \mathbf{W} \mathbf{V}^{-1} \boldsymbol{g}.$$

For all elements we may precompute $(\mathbf{V}^{-1})^{\mathrm{T}} \mathbf{W} \mathbf{V}^{-1}$ in a preprocessing stage, storing only the upper half of the operator due to symmetry. In the particularly important case where D is a straightsided tetrahedron, i.e., J is a constant, the orthonormality of ψ_α implies that $\mathbf{W} = J\mathbf{I}$, where $\mathbf{I}$ represents the identity matrix. Hence, through a simple linear scaling one recovers the weights for all tetrahedra with planar faces. For the general case where $J(\boldsymbol{\xi})$ is nonconstant, the entries of $\mathbf{W}$ are computed exactly through overintegration by product rules based on Legendre Gauss quadratures [54].

A final key operation needed for the implementation of the scheme is surface integration, i.e.,

$$(f_N, g_N)_{\delta\mathsf{D}} = \oint_{\delta\mathsf{I}} f_N(\boldsymbol{\xi}) g_N(\boldsymbol{\xi}) J(\boldsymbol{\xi})\, d\boldsymbol{\xi},$$

where $J(\boldsymbol{\xi})$ refers to the surface Jacobian only. While one could proceed as for the volume integral discussed above, it is more natural to exploit the uniqueness and completeness of the Lagrange interpolation. To illustrate the procedure, let us restrict attention to one of the faces, $\zeta = -1$, and term those $N_n^{\mathrm{f}} = \frac{1}{2}(n+1)(n+2)$ nodes positioned at that face for $\boldsymbol{\xi}^{\mathrm{f}}$. Clearly, using the exact same procedure as for the three-dimensional Lagrange polynomial

discussed above, we can compute a two-dimensional Lagrange polynomial, $l_j^{\mathrm{f}}(\xi,\eta)$ based on $\boldsymbol{\xi}^{\mathrm{f}}$. As for $L_j(\boldsymbol{\xi})$, we can recover l_j^{f} as the solution to the dual problem

$$(\mathbf{V}^{\mathrm{f}})^{\mathrm{T}}\boldsymbol{l}^{\mathrm{f}} = \boldsymbol{p}^{\mathrm{f}},$$

where the entries of the Vandermonde matrix are

$$\mathbf{V}_{ij}^{\mathrm{f}} = p_j^{\mathrm{f}}(\boldsymbol{\xi}_i^{\mathrm{f}}).$$

The proper basis to use is the two-dimensional version of Eq. (A.1) given directly as $p_j^{\mathrm{f}}(\xi,\eta) = \psi_{ij0}(\xi,\eta,-1)$. This allows us to proceed exactly as for the volume integration and express the integration over the face as

$$\int_{\text{face f}} f_N(\xi,\eta,-1)g_N(\xi,\eta,-1)J(\xi,\eta,-1)\,d\xi\,d\eta = ((\mathbf{V}^{\mathrm{f}})^{-1}\boldsymbol{f}^{\mathrm{f}})^{\mathrm{T}}\mathbf{W}^{\mathrm{f}}(\mathbf{V}^{\mathrm{f}})^{-1}\boldsymbol{g}^{\mathrm{f}},$$

where $\boldsymbol{f}^{\mathrm{f}} = [f_N(\boldsymbol{\xi}_0^{\mathrm{f}}),\ldots,f_N(\boldsymbol{\xi}_{N_n^{\mathrm{f}}}^{\mathrm{f}})]^{\mathrm{T}}$ is the trace of f_N at the face. A similar definition is used for $\boldsymbol{g}^{\mathrm{f}}$. The matrix of surface weights are given as

$$\mathbf{W}_{ij}^{\mathrm{f}} = \int_{\text{face f}} \psi_i(\xi,\eta,-1)\psi_j(\xi,\eta,-1)J(\xi,\eta,-1)\,d\xi\,d\eta.$$

In the important special case where the face is planar and has straight edges, orthonormality of the polynomials immediately implies that $\mathbf{W}^{\mathrm{f}} = J^{\mathrm{f}}\mathbf{I}$ as for the volume case. For the general case we shall use a cubature rule [55–57] of sufficiently high order to evaluate the inner product; i.e., we need to interpolate the polynomials, f_N and g_N, onto the M cubature nodes, $\boldsymbol{\xi}^{\mathrm{d,cub}}$, situated at the face. This is done by the introduction of the interpolation operator

$$\mathbf{H} = \mathbf{P}^{\mathrm{T}}(\mathbf{V}^{\mathrm{f}})^{-1}, \quad \mathbf{P}_{ij} = p_i^{\mathrm{f}}\big(\boldsymbol{\xi}_j^{\mathrm{f,cub}}\big);$$

i.e., $\mathbf{P}$ is an $N_n^{\mathrm{f}} \times M$ operator. The evaluation of the inner product is then accomplished as

$$\int_{\text{face f}} f_N(\xi,\eta,-1)g_N(\xi,\eta,-1)J(\xi,\eta,-1)\,d\xi\,d\eta = (\boldsymbol{f}^{\mathrm{f}})^{\mathrm{T}}\mathbf{H}^{\mathrm{T}}\mathbf{W}\mathbf{H}\boldsymbol{g}^{\mathrm{f}},$$

where the entries of the diagonal $M \times M$ matrix of weights are given as

$$\mathbf{W}_{ii} = w_i \sum_{k=0}^{N_n^{\mathrm{f}}} \mathbf{H}_{ik} J\big(\boldsymbol{\xi}_k^{\mathrm{f}}\big),$$

containing the weights w_i of the cubature as well as the interpolation of the transformation Jacobian of the curvilinear face. While this formulation leads to the most compact scheme it proves advantageous to operate directly on the values at the cubature nodes as they do not include the edges and vertices; i.e., we can establish a clean face-based connection between elements without considering the multiplicity of solutions at vertices and the added complexity this introduces for the implementation and performance.

It is important to realize that all the operators introduced in the above can be initialized during a preprocessing phase. Furthermore, it is worth recalling the discussion in Section 3.1 in which we found that any two straightfaced tetrahedra are connected through a linear transformation. Hence, for any straightfaced D we can form any of the operators discussed in the above directly by a linear scaling of hard-coded template operators defined on I. This saves not only preprocessing time but also reduces the required storage space very substantially.

ACKNOWLEDGMENTS

The authors acknowledge the partial support of AFOSR/DARPA under Contract F49620-1-0426 and F33615-01-C-1866 and by the National Aeronautics and Space Administration under NASA Contract NAS1-97046 while the authors were in residence at ICASE, NASA Langley Research Center, Hampton, VA 23681-2199. J.S.H. also acknowledge support by the Alfred P. Sloan Foundation as a Sloan Research Fellow. The authors finally extend their appreciation to Professor D. Gottlieb and Dr. A. Ditkowski, Brown University, for many fruitful discussions.

REFERENCES

1. H. O. Kreiss and J. Oliger, Comparison of accurate methods for the integration of hyperbolic problems, *Tellus* **24**, 199–215 (1972).
2. A. T. Patera, A spectral element method for fluid mechanics: Laminar flow in a channel expansion, *J. Comput. Phys.* **54**, 468–488 (1984).
3. K. Z. Korczak and A. T. Patera, An isoparametric spectral element method for solution of the Navier–Stokes equations in complex geometries, *J. Comput. Phys.* **62**, 361–382 (1986).
4. C. Canuto, M. Y. Hussaini, A. Quarteroni, and T. A. Zang, *Spectral methods in fluid dynamics*, Springer Series in Computational Physics (Springer-Verlag, New York, 1988).
5. D. A. Kopriva, A spectral multidomain method for the solution of hyperbolic systems, *Appl. Numer. Math.* **2**, 221–241 (1986).
6. D. A. Kopriva, Computation of hyperbolic equations on complicated domains with patched and overset Chebyshev grids, *SIAM J. Sci. Stat. Comput.* **10**, 120–132 (1989).
7. A. Quarteroni, Domain decomposition methods for systems of conservation laws: Spectral collocation approximations, *SIAM J. Sci. Stat. Comput.* **11**, 1029–1052 (1990).
8. D. A. Kopriva, Multidomain Spectral Solution of the Euler gas-dynamics equations, *J. Comput. Phys.* **96**, 428–450 (1991).
9. J. G. Giannakouros and G. E. Karniadakis, Spectral element-FCT method for the compressible Euler equations, *J. Comput. Phys.* **115**, 65–85 (1994).
10. J. S. Hesthaven, A stable penalty method for the compressible Navier–Stokes equations. III. Multidimensional domain decomposition schemes, *SIAM J. Sci. Comput.* **20**, 62–93 (1999).
11. B. Yang, D. Gottlieb, and J. S. Hesthaven, Spectral simulations of electromagnetic wave scattering, *J. Comput. Phys.* **134**, 216–230 (1997).
12. B. Yang and J. S. Hesthaven, A pseudospectral method for time-domain computation of electromagnetic scattering by bodies of revolution, *IEEE Trans. Ante. Prop.* **47**, 132–141 (1999).
13. J. S. Hesthaven, P. G. Dinesen, and J. P. Lynov, Spectral collocation time-domain modeling of diffractive optical elements, *J. Comput. Phys.* **155**, 287–306 (1999).
14. D. A. Kopriva, S. L. Woodruff, and M. Y. Hussaini, Discontinuous spectral element approximation of Maxwell's equations. In *Discontinuous Galerkin Methods: Theory, Computation and Applications*, edited by B. Cockburn, G. E. Karniadakis, and C. W. Shu, Lecture Notes in Computational Science and Engineering (Springer-Verlag, New York, 2000), Vol. 11, pp. 355–362.

15. B. Yang and J. S. Hesthaven, Multidomain pseudospectral computation of Maxwells equations in 3-D general curvilinear coordinates, *Appl. Numer. Math.* **33**, 281–289 (2000).
16. S. J. Sherwin and G. E. Karniadakis, A new triangular and tetrahedral basis for high-order finite element methods, *Int. J. Numer. Methods* **38**, 3775–3802 (1995).
17. S. J. Sherwin and G. E. Karniadakis, Tetrahedral hp finite elements: Algorithms and flow simulations. *J. Comput. Phys.* **124**, 14–45 (1996).
18. I. Lomtev, C. B. Quillen, and G. E. Karniadakis, Spectral/hp methods for viscous compressible flows on unstructured 2D meshes, *J. Comput. Phys.* **144**, 325–357 (1998).
19. T. Warburton and G. E. Karniadakis, A discontinuous Galerkin method for the viscous MHD equations, *J. Comput. Phys.* **152**, 608–641 (1999).
20. T. Warburton, I. Lomtev, Y. Du, S. Sherwin, and G. E. Karniadakis, Galerkin and discontinuous Galerkin spectral/hp methods, *Comput. Methods Appl. Mech. Engrg.* **175**, 343–359 (1999).
21. G. E. Karniadakis and S. J. Sherwin, *Spectral/hp Element Methods for CFD*, Numerical Mathematics and Scientific Computation (Clarendon Press, Oxford, 1999).
22. H. Atkins and C. W. Shu, Quadrature-free implementation of discontinuous Galerkin methods for hyperbolic equations, *AIAA J.* **36**, 775–782 (1998).
23. B. Cockburn, G. E. Karniadakis, and C.-W. Shu (Eds.), *Discontinuous Galerkin Methods: Theory, Computation and Applications*, Lecture Notes in Computational Science and Engineering (Springer-Verlag, New York, 2000), Vol. 11.
24. I. Babuška and M. Suri, The hp-Version of the finite element method with quasiuniform meshes, *Math. Model. Numer. Anal.* **21**, 199–238 (1987).
25. C. Schwab, *p- and hp-Finite Element Methods. Theory and Applications in Solid and Fluid Mechanics*, Numerical Mathematics and Scientific Computation (Clarendon Press, Oxford, 1998).
26. J. S. Hesthaven, Spectral penalty methods, *Appl. Numer. Math.* **33**, 23–41 (2000).
27. M. H. Carpenter and D. Gottlieb, Spectral methods on arbitrary grids, *J. Comput. Phys.* **129**, 74–86 (1996).
28. J. S. Hesthaven and D. Gottlieb, Stable spectral methods for conservation laws on triangles with unstructured grids, *Comput. Methods Appl. Mech. Engrg.* **175**, 361–381 (1999).
29. J. S. Hesthaven and C. H. Teng, Stable spectral methods on tetrahedral elements, *SIAM J. Sci. Comput.* **21**, 2352–2380 (2000).
30. T. Warburton and J. S. Hesthaven, Nodal high-order methods on unstructured grids. II. Systems of nonlinear conservation laws, *J. Comput. Phys.*, in preparation.
31. Q. Chen and I. Babuška, The optimal symmetrical points for polynomial interpolation of real functions in a tetrahedron, *Comput. Methods Appl. Mech. Engrg.* **137**, 89–94 (1996).
32. A. H. Mohammadian, V. Shankar, and W. F. Hall, Computation of electromagnetic scattering and radiation using a time-domain finite-volume discretization procedure, *Comput. Phys. Comm.* **68**, 175–196 (1991).
33. I. Babuška and A. K. Aziz, On the angle condition in the finite element method, *SIAM J. Numer. Anal.* **13**, 214–226 (1976).
34. P. G. Ciarlet, *The Finite Element Method for Elliptic Problems* (North-Holland, The Netherlands, 1978).
35. A. Ditkowski, *Bounded-Error Finite Difference Schemes for Initial Boundary Value Problems on Complex Domains*, Ph.D. thesis (Department of Applied Mathematics, School of Mathematical Sciences, Tel-Aviv University, Tel-Aviv, Israel, 1997).
36. A. Ditkowski, K. Dridi, and J. S. Hesthaven, Convergent Cartesian grid methods for Maxwells equations in complex geometries, *J. Comput. Phys.* **170**, 39–80 (2001).
37. E. Suli, C. Schwab, and P. Houston, hp-DGFEM for partial differential equations with nonnegative characteristic form, in *Discontinuous Galerkin Methods. Theory, Computation and Applications,* edited by B. Cockburn, G. E. Karniadakis, and C. W. Shu, Lecture Notes in Computational Science and Engineering (Springer-Verlag, Berlin, 2000), Vol. 11, pp. 221–230.
38. M. H. Carpenter and C. A. Kennedy, Fourth order 2N-storage Runge-Kutta scheme, NASA-TM-109112 (NASA Langley Research Center, VA, 1994).
39. S. Abarbanel and D. Gottlieb, On the construction and analysis of absorbing layers in CEM, *Appl. Numer. Math.* **27**, 331–340 (1998).

40. A. Taflove, Computational Electrodynamics: *The Finite-Difference Time-Domain Method* (Artech House, Boston, 1995).
41. J. J. Bowman, T. B. A. Senior, and P. L. Ushlenghi (Eds.), *Electromagnetic and Acoustic Scattering by Simple Shapes* (North-Holland, Amsterdam, 1969).
42. P. W. Barber and S. C. Hill, *Light Scattering by Particles: Computational Methods* (World Scientific, Singapore, 1990).
43. J. L. Volakis, Benchmark plate radar targets for the validation of computational electromagnetics programs, *IEEE Ant. Prop. Mag.* **34**, 52–56 (1992).
44. J. Dongarra, J. Du Croz, I. Duff, and S. Hammerling, A set of level 3 basic linear algebra subprograms (BLAS), available at http://www.netlib.org/blas/blas3-paper.ps.
45. M. Snir, S. Otto, S. Huss-Lederman, D. Walker, and J. Dongarra, *MPI: The Complete Reference* (MIT Press, Cambridge, MA, 1996).
46. G. Karypis and V. Kumar, Multilevel k-way partitioning scheme for irregular graphs, *J. Para. Distrib. Comput.* **48**, 96–129 (1998).
47. Y. Saad, *Iterative Methods for Sparse Linear Systems*, 2nd ed. (2000), available at http://www-users.cs.umn.edu/∼saad/books.html.
48. C. C. Douglas, G. Haase, J. Hu, M. Kowarschik, U. Rüde, and C. Weiss, Portable memory hierarchy techniques for PDE solvers: Part II, *SIAM News* **33** (2000).
49. R. Biswas, K. D. Devine, and J. Flaherty, Parallel, Adaptive finite element methods for conservation laws, *Appl. Numer. Math.* **14**, 255–283 (1994).
50. J. Proriol, Sur une Famille de Polynomes à deux Variables Orthogonaux dans un Triangle, *C. R. Acad. Sci. Paris* **257**, 2459–2461 (1957).
51. T. Koornwinder, Two-variable analogues of the classical orthogonal polynomials, in *Theory and Application of Special Functions*, edited by R. A. Askey (Academic Press, New York, 1975), pp. 435–495.
52. M. Dubiner, Spectral methods on triangles and other domains, *J. Sci. Comput.* **6**, 345–390, 1991.
53. G. Szegö, *Orthogonal Polynomials.* Colloquium Publications (Am. Math. Soc., Providence, RI, 1939), Vol. 23.
54. P. J. Davis and P. Rabinowitz, *Methods of Numerical Integration.* Computer Science and Applied Mathematics (Academic Press, New York, 1975).
55. A. H. Stroud, *Approximate Calculation of Multiple Integrals* (Prentice–Hall, New Jersey, 1971).
56. R. Cools and P. Rabinowitz, Monomial cubature rules since Stroud: A compilation, *J. Comput. Appl. Math.* **48**, 309–326 (1993).
57. R. Cools, Monomial cubature rules since Stroud: A compilation—Part 2, *J. Comput. Appl. Math.* **112**, 21–27 (1999).